

**Alexandre FRANÇOIS**

LACITO-CNRS, Paris;  
Australian National University

# **Structures et dynamiques des systèmes linguistiques**

**Description, documentation,  
comparaison des langues océaniennes**

---

Travaux rassemblés en vue de  
l'Habilitation à Diriger des Recherches

Volume II

## **PUBLICATIONS SCIENTIFIQUES**

Articles de revues à comité de lecture  
Chapitres d'ouvrages

2014

Institut National des Langues et Civilisations Orientales, Paris

Alexandre FRANÇOIS  
*Habilitation à Diriger des Recherches*

Publications scientifiques :  
Articles et chapitres

---

## **Chapitres d'ouvrages scientifiques**

---

## Chapitres d'ouvrages scientifiques

- (p14) François, Alexandre. 2000 a. **Vowel shifting and cloning in Motlav : historical explanation vs. formal description.** In Marian Klamer (ed.), *Proceedings of AFLA7 (The Seventh Meeting of Austronesian Formal Linguistics Association)*. Amsterdam: Vrije Universiteit. Pp.49-68.
- (p15) François, Alexandre. 2004 a. **Chains of freedom: Constraints and creativity in the macro-verb strategies of Mwotlap.** In I. Bril & F. Ozanne-Rivierre (eds), *Complex predicates in Oceanic languages: Studies in the dynamics of binding and boundness. Empirical Approaches to Language Typology*. Berlin: Mouton de Gruyter. Pp.107-143.
- (p16) François, Alexandre. 2004 d. **Diversité des prédicts non verbaux dans quelques langues océaniennes.** In J. François & I. Behr (eds), *Les constituants prédictifs et la diversité des langues: Actes de la Journée de la Société de Linguistique de Paris*. Mémoires de la Société de Linguistique de Paris. Louvain: Peeters. Pp.179-197.
- (p17) François, Alexandre. 2006. **Serial verb constructions in Mwotlap.** In R.M.W. Dixon & A. Aikhenvald (eds), *Serial Verb Constructions: A cross-linguistic typology. Explorations in Linguistic Typology*. Oxford: Oxford University Press. Pp.223-238.
- (p18) François, Alexandre. 2007. **Noun articles in Torres and Banks languages: Conservation and innovation.** In J. Lynch, J. Siegel & D. Eades (eds), *Language Description, History and Development: Linguistic Indulgence in Memory of Terry Crowley*. Creole Language Library, 30. New York: John Benjamins. Pp.313-326.
- (p19) François, Alexandre. 2008 a. **Semantic maps and the typology of colexification: Intertwining polysemous networks across languages.** In M. Vanhove (ed.), *From Polysemy to Semantic change: Towards a Typology of Lexical Semantic Associations*. Studies in Language Companion Series, 106. New York, Amsterdam: Benjamins. Pp.163-215.
- (p20) François, Alexandre. 2009 a. **The languages of Vanikoro: Three lexicons and one grammar.** In B. Evans (ed.), *Discovering history through language: Papers in honour of Malcolm Ross*. Pacific Linguistics, 605. Canberra: Australian National University. Pp.103-126.
- (p21) François, Alexandre. 2009 b. **Verbal aspect and personal pronouns: The history of aorist markers in north Vanuatu.** In A. Pawley & A. Adelaar (eds), *Austronesian historical linguistics and culture history: A festschrift for Bob Blust*. Pacific Linguistics, 601. Canberra: Australian National University. Pp.179-195.
- (p22) François, Alexandre. 2010 c. **Pragmatic demotion and clause dependency: On two atypical subordinating strategies in Lo-Toga and Hiw (Torres, Vanuatu).** In I. Bril (ed.), *Clause hierarchy and Clause linking: the Syntax and pragmatics interface*. Studies in Language Companion Series 121. Amsterdam, New York: Benjamins. Pp.499-548.
- (p23) François, Alexandre. 2013 a. **Shadows of bygone lives: The histories of spiritual words in northern Vanuatu.** In R. Mailhammer (ed.), *Lexical and structural etymology: Beyond word histories*. Studies in Language Change. Berlin: DeGruyter Mouton. Pp.185-244.
- (p24) François, Alexandre. (ss presse a). **Temperature terms in Vanuatu languages.** In Maria Koptjevskaja Tamm (ed.), *Linguistics of Temperature*. Amsterdam, New York: Benjamins.
- (p25) François, Alexandre. (ss presse b). **The ins and outs of up and down: Disentangling the nine geocentric space systems of Torres and Banks languages.** In A. François; S. Lacrampe; S. Schnell & M. Franjieh (eds), *The Languages of Vanuatu: Unity and Diversity*. Studies in the Languages of Island Melanesia. Canberra: Asia Pacific Linguistics Open Access. 74 pp.

- (p26) Kalyan, Siva, & Alexandre François. (ss presse).
- Freeing the Comparative Method from the tree model: A framework for Historical Glottometry.** In R. Kikusawa & L. A. Reid (eds), *Let's talk about trees: Tackling Problems in Representing Phylogenetic Relationships among Languages*. Senri Ethnological Studies. Osaka: National Museum of Ethnology.
- (p27) François, Alexandre. (ss presse c).
- Trees, Waves and Linkages: Models of Language Diversification.** In C. Bowern & B. Evans (eds), *The Routledge Handbook of Historical Linguistics*. New York: Routledge.

# Vowel shifting and cloning in Mwotlap: historical explanation vs. formal description

Alexandre FRANÇOIS

LACITO-CNRS, Paris

Université Paris-Sorbonne

## Abstract

In his first account of Mwotlap grammar, Codrington (1885), followed by Kasarhérou (1962), described the phonological rule of ‘vowel shifting’: a number of prefixes change their vowel into that of the next lexeme, e.g. *na-* + *vōy* > *nō-vōy* (‘volcano’). We first show that this rule only accounts for half of the lexicon, namely CV- roots, whereas for CCV-roots no change occurs, leaving an unvarying vowel instead.

We then discuss a diachronic hypothesis in order to account for these two distinct sets of lexemes: stress rules in former stages of the language first brought about a morphological alternation between two kinds of roots (CV- vs. CCV-). Eventually, this alternation was reinterpreted as reflecting a phonological constraint just emerging from inherited forms; as a consequence, the whole system is currently undergoing some standardization.

Finally, we provide a synchronic reanalysis of these recently emerged rules, thanks to newly defined theoretical tools. For instance, our choice of a multi-linear approach allows us to take into consideration the notion of ‘floating phoneme’, in order to account for the behavior of these prefixes; and the same notion appears to be efficient in analyzing ‘vowel transfer’ (e.g. *na-* + *hinag* > *ni-hnag*), another phonological rule of the same language.

---

Mwotlap is an Oceanic language spoken in northern Vanuatu (Banks Is.), on the island of Mwotlap/Motalava, by approximately 1800 speakers<sup>1</sup>. It is geographically and historically close to Mota, a more conservative language which was first described in detail by Reverend Codrington (1885; 1896); this author also gave a first account of the grammar of Mwotlap (1885: 311-323).

Contrary to Mota, the morphology of Mwotlap is quite difficult, and seems to obey complex segmental and phonotactic rules, leaving bundles of lexical exceptions. Our aim is to show that this apparent complexity may in fact be reduced to a small set of rules and constraints, which mostly affect the quality and position of vowels within the word.

Yet, such a simplification of the analysis requires some strengthening or redefinition of our theoretical tools, which form the framework of our analysis: we will then come across such notions as syllabic template, floating vowels, and hierarchy of cognitive operations.

---

<sup>1</sup> An earlier but more detailed version of this article was published (François 1999), relating the morphology of vowels in Mwotlap to dictionary-making issues.

## 1. Phonological outline of Mwotlap

### 1.1. Consonants

Mwotlap contains 16 distinct consonants, which may appear in any position of the linguistic chain – i.e. syllable-initial or final. Their phonological value is given below, together with their spelling when it is not obvious.

	<i>labiovelar</i>	<i>bilabial</i>	<i>alveol.</i>	<i>velar</i>	<i>glottal</i>
<i>voiceless stops</i>	<sup>k</sup> p <sup>w</sup>  q		t	k	
<i>prenas. voiced stops</i>		<sup>m</sup> b  b	<sup>n</sup> d  d		
<i>fricatives</i>		v	s	γ  g	h
<i>nasals</i>	<sup>ŋ</sup> m <sup>w</sup>  m̄	m	n	ŋ  n̄	
<i>lateral</i>			l		
<i>glides</i>	w		j  y		

This information is being given here for the reader's convenience; but it may be useful to underline the fact that, synchronically speaking, there is no morphological rule in Mwotlap involving the quality or position of consonants, such as assimilation.

Let us only underline the fact that the two glides /w/ and /y/ strictly behave as consonants in this language: they never form diphthongs, and follow the same rules as other Cs when filling in the syllable template. Note also that the phoneme /v/ surfaces as [p] at the end of a syllable, and this phonetic allophone<sup>2</sup> is spelt *p* in the orthography: e.g. *na-pnō* ‘country’ corresponds phonologically to /na-vnō/.

### 1.2. Vowels

However complex the rules involving vowels may be, the vowel system in itself is quite simple, since it contains no diphthongs, nor long or nasalized vowels, but instead seven oral, short vowels, represented below.

i	u
I  e̥	U  o̥
e	o
a	

Codrington (1885: 311) only mentioned six vowels, while other analyses (Crowley forthc.; Kasarhérou 1962) proposed a seven-element system, but with different vowel qualities. In another article (François 1999: 443), we have claimed that the two mid-high vowels should be better described as being [-ATR], both on phonetic and phonemic grounds: the [ATR] feature is involved in the morphological process of vowel harmonization, which we will not deal with here.

<sup>2</sup> In his grammar sketch of Mwotlap, Terry Crowley (forthc.) considers the phoneme /p/, although it shows no phonological contrast with /v/.

### 1.3. Phonotactics

Mwotlap only accepts one type of syllabic pattern, i.e. |(C)V(C)|. Since both consonants are optional here, the phonology of Mwotlap actually contains four patterns: V, CV, VC, CVC; but the essential point here is to underline that there is no allowance for consonant clusters within a syllable, and that -C<sub>1</sub>C<sub>2</sub>- sequences may only show at the syllable boundary.

As a consequence, a CCV- morpheme will only surface as such if its first consonant can attach to a preceding open-syllable (CV-) prefix. For example, the radical *mtiy* ‘sleep’ will stay unchanged in *ni-mtiy* ‘(s/he) sleeps’, because it may be segmented into two CVC syllables |*nim/tiy*|; whereas an unprefixed form, as the one for imperative ‘Sleep!’, will have to undergo a process of vowel epenthesis, with the vowel of the radical cloning between the first consonant cluster: \**mtiy* → *M<sub>i</sub>tiy!*, segmented |*mi/tiy*|. We won’t detail this rule here<sup>3</sup>, but understanding it gives a first notion of the kind of phonotactic constraints defining Mwotlap morpho-phonology.

We can characterize the notion of phonological word in Mwotlap, as a segmental unit including one or several morphemes, which are together bound to a strict [CVC]... [CVC] pattern. Incidentally, this definition makes it possible to distinguish *prefixes* from *clitics* on phonological grounds: a prefix is integrated to the phonological word, and as such takes its position inside the syllabic pattern thus defined – e.g. *le-* is a prefix in /*le-p|nō*/ ‘in the country’, without epenthesis. Whereas a clitic only integrates into the *stress-defined* word<sup>4</sup>, but never into the *phonotactic* word: e.g. *ne* is a clitic in /*ne||vō|nō*/ ‘of the country’, because it remains without the word boundary defined on phonotactic grounds.

## 2. Distribution of vowels within the word

The four main phonological rules involving vowels in Mwotlap are designated below:

VOWEL HARMONIZATION:	<i>iplu-k</i> ‘my fellow’ → <i>ēplō-n</i> ‘his fellow’;
VOWEL EPENTHESIS:	<i>mtig</i> → <i>mitig</i> ‘coconut’;
VOWEL SHIFTING:	<i>na- + wōl</i> → <i>nō-wōl</i> ‘moon’;
VOWEL TRANSFER:	<i>na- + hinag</i> → <i>ni-hnag</i> ‘yam’.

The first two rules have already been alluded to, and won’t be detailed here; this paper will deal mostly with vowel shifting, and will eventually present the operation of vowel transfer (§5).

### 2.1. Evidence for vowel shifting

The phenomenon of vowel shifting was first acknowledged by Codrington (1885: 311), and focused upon by Jacqueline Kasarhérou (de la Fontinelle), in her short presentation ‘Les changements vocaliques de trois préfixes en motlav’ (1962). Both authors presented this rule as involving a series of prefixes consisting of one single consonant (C-): e.g. the noun article *n-*, the verbal prefix *n-* (considered by Kasarhérou to be the same as the article), the perfect marker *m-*, would either surface as such (C-) before a root beginning with

---

<sup>3</sup> See François (1999: 475-479).

<sup>4</sup> Mwotlap words receive stress on their last syllable: see fn.9.

a V, or else would borrow their vowel from the root, through a cloning / shifting operation:

**Table 1 - Rules for vowel shifting, with noun article n-**

radical	*n- + radic.	radical	article+rad.	meaning
V <sub>i</sub> -	n-V <sub>i</sub> -	ulsi	n-ulsi	summit
CV <sub>i</sub> -	nV <sub>i</sub> - CV <sub>i</sub> -	bē	nē-bē	water
CCV <sub>i</sub> -	nV <sub>i</sub> -C CV <sub>i</sub> -	qti	ni-qti	head

Further evidence of this phenomenon include the following items, among thousands. The first rows involve the noun article (supposedly *n*-), the last one uses a numeralizer of the form *v*-:

**Table 2 - Sample evidence for vowel shifting**

<i>ni-git</i>	louse	<i>nē-sēm</i>	money	<i>ne-qet</i>	taro
<i>no-gom</i>	disease	<i>nō-vōy</i>	volcano	<i>nu-kumay</i>	sweet potato
<i>na-lan̄</i>	(a) fly	<i>ni-qti-k</i>	my head	<i>nē-qtē-n</i>	his head
<i>vō-yō</i>	two	<i>ve-vet</i>	four	<i>vē-vēh</i>	how many ?

This insertion of a shifted vowel can be analyzed as an epenthesis rule to avoid consonant clusters at the beginning of a word. This is exactly what happens with another of the phonological rules in Mwotlap, which we have labeled ‘vowel epenthesis’: a CCV- root like *mtig* ‘coconut’ needs to copy its vowel into the consonant cluster, to avoid it at the beginning of a word (*mtig* → *mitig*). This recalls also what happens in Kalam, a non-Austronesian language from Papua New Guinea (Pawley 1993: 91):

In the context C\_CVC, the release vowel may be a very short, unstressed near copy of V or a short, unstressed central or high central vowel, e.g., *mlep* ‘dry’ is [melep] or [miɻep].

In Mwotlap, no central vowel is heard, and the epenthetic vowel is always clearly a clone of the following V. These prefixes “seem not to have a proper vowel, but instead to borrow it from the radical” (Kasarhérou 1962). This evidence should suffice for positing suffixes of the form C-, i.e. with a single consonant, and no specific vowel.

## 2.2. Exceptions to vowel shifting

Despite the obviousness of the latter analysis, we have to face a quantity of irregular forms, whose prefix vowel is different from that of the radical. A few of them are presented in *Table 3*, where irregular forms are underlined:

**Table 3 - Some exceptions to vowel shifting**

+ article	meaning	+ locative	meaning
<i>na-naw</i>	sea	<u><i>le-naw</i></u>	in the sea
<i>na-s̄mal</i>	rain	<u><i>le-s̄mal</i></u>	in the rain
<u><i>na-he-k</i></u>	my name	<i>le-he-k</i>	in my name
<u><i>na-gmel</i></u>	men's house	<i>le-gmel</i>	in the men's house
<u><i>na-lo</i></u>	sun	<u><i>le-lo</i></u>	in the sun
<u><i>na-pnō</i></u>	country, island	<u><i>le-pnō</i></u>	in the country

We have chosen only a few radicals among those which show at least one irregular form; each radical is presented here with two different prefixes, the noun article *n-* and the locative preposition *l-*. A regular pair for both prefixes would be *nō-vōy* ‘volcano’ – *lō-vōy* ‘on the volcano’. *Table 3* does show regular forms too, like *na-naw* or *le-gmel*; but the rest is not, and falls among the ‘exceptions’ pointed at by both Codrington and Kasarhérou.

The trouble is, such ‘exceptions’ are not few, but represent up to 50% of the data. Such observation obviously needs to be accounted for, unless the rule for vowel shifting appears to be just a weak random trend in Mwotlap phonology. In fact, another sort of regularity appears in *Table 3*, provided it is not read in rows, but in columns. For a given prefix, all the irregular forms will show the same vowel; e.g. every noun preceded by the article *n-*, will have as a first vowel either a copy of the radical vowel (e.g. *nō-vōy*), or the vowel /a/ (e.g. *na-gmel*, *na-lo*); as for the locative *l-*, the ‘default vowel’, as it may be called, is an /e/ (e.g. *le-naw*, *le-lo*).

### 2.3. *Shifting prefixes vs. unvarying prefixes*

Evidently, Mwotlap does not have just one neutral vowel which could be assigned at the system-level, like Indonesian Pendau does with /o/ (Quick 2000); instead, each shifting prefix is provided with its own ‘default vowel’. The quality of this vowel is assigned to the prefix in the lexicon, and cannot be predicted through any phonological rule: it is thus necessary to represent it in the citation form of each morpheme. From now on, we will choose to speak about the noun article *na-* or the locative preposition *le-*, instead of *n-* and *l-* respectively, since this is the only way to know what their default vowel is. This is a first difference with former accounts which were given of Mwotlap phonology, since we now consider shifting prefixes to be of the form CV-, with both C and V being specific.

Nevertheless, one improvement has to be made in our representation of these CV-prefixes. Talking about an article *na-* would make it necessary to posit an extra rule of deletion / assimilation for this /a/ before certain radicals, e.g. *na- + qōn* → *nō-qōn* ‘night’. This rule would have a form like

$$< C_0 V_{0-} + C_1(C_2) V_1 C_3 \dots \rightarrow C_0 \boxed{V_1} - C_1(C_2) V_1 C_3 \dots >$$

However, an argument against such a generalization is given by other prefixes of the form (C)V-, which never change their vowel:

- verb prefix *ni-* (3<sup>rd</sup> singular + present): e.g. *ni-in* ‘drinks’, *ni-et* ‘sees’, *ni-van* ‘goes’, *ni-gen* ‘eats’, *ni-qōn* ‘(it) becomes night’;
- locative prefix *a-*, showing in several place names: *a-Pnōlap* ‘Vanua Lava’, *a-Gō* ‘Gaua’, *a-Mēw* ‘Maewo’, *a-Msēn* ‘Mosina’, *a-Mōt* ‘Mota’.

As a consequence, vowel shifting in Mwotlap cannot be described by the general rule stated above. Some CV- prefixes are unvarying, while others do show a ‘weak’ vowel, which is likely to assimilate to the vowel of the radical. This is why the best representation is to oppose these two kinds of CV- prefixes, typographically, by contrasting:

- prefixes of the form /CV-/ which never shift their vowel, e.g. *ni-* (3SG:PST); they will be called ‘unvarying prefixes’;
- prefixes of the form /CV-/, which are sensitive to vowel shifting, e.g. *nA-* (noun article); they are the ‘shifting prefixes’.

The uppercase letter codes for the special status of this ‘default vowel’, i.e. a vowel which sometimes surfaces and sometimes does not; the exact theoretical status of this vowel will be the main issue of §4. Mwotlap has eight shifting prefixes, which are listed below:

**Table 4 - The eight vowel-shifting prefixes of Mwotlap**

FORM	MEANING	PREFIXED TO
<i>nA-</i>	Article ('a / the')	nouns
<i>bE-</i>	Purposive ('for')	
<i>lE-</i>	Locative ('in')	
<i>mE-</i>	Perfect	verbs, adjectives, (nouns)
<i>nE-</i>	Stative	
<i>tE-</i>	Future	
<i>tE-</i>	Ablative	locatives
<i>vE-</i>	Numerilizer	numbers

Incidentally, notice that some of these prefixes can combine. This is mainly the case for *tE-* ('Ablative'), which is prefixed to locative words, be they directly locative –e.g. *to-Motlap* ‘from Mwotlap’–, or derived by means of the preposition *lE-* ('Locative'), thus bringing about a sequence *tE-lE-*. In this case, the preposition which is adjacent to the root (*lE-*) is sensitive to vowel shifting, and its vowel is always reproduced on the first prefix (*tE-*):

- |                        |   |                  |                    |
|------------------------|---|------------------|--------------------|
| <i>tE- + lE- + ēm̄</i> | → | <i>tē-l-ēm̄</i>  | 'from the house'   |
| <i>tE- + lE- + vōy</i> | → | <i>tō-lō-vōy</i> | 'from the volcano' |
| <i>tE- + lE- + naw</i> | → | <i>te-le-naw</i> | 'from the sea'     |

At this stage of the presentation, it is impossible to know whether the vowel /e/ on the first prefix (*te-*) is an instance of its own ‘default vowel’ –which would mean that no vowel shifting occurred at all–, or if it is a copy of the following vowel –which means that vowel shifting only affects the second prefix (*lE-*). Further demonstration will show the second assumption to be correct.

#### 2.4. Permeable vs. blocking lexemes

One of the conditions for vowel shifting, obviously, is that the prefix must belong to the limited list of shifting prefixes (*Table 4*). But what about the radical? Is there any restriction on the set of radicals that can shift their vowel, and is that restriction systematically predictable?

*Table 3* above had already provided a set of exceptions to the vowel-shifting process. However, only the underlined forms (e.g. *le-naw*) were supposed to be irregular, while other forms, including for the same radical (e.g. *na-naw*), were considered to follow the rules. But now that we know the underlying form of each prefix, it becomes ambiguous whether the first /a/ in *na-naw* is really a clone of the vowel in the radical (*naw*), or if it is the default vowel of the noun article (*nA-*). Now, there is strong evidence pointing at the latter solution: *whenever a given lexical root blocks the vowel-shifting process with one prefix, it does so with all other prefixes*. Conversely, whenever a lexeme shifts its vowel to one prefix, it will do so with any other (shifting) prefix. *Table 5* gives further evidence of this claim:

**Table 5 - Permeable vs. blocking lexemes**

	trans.	bare root	article <i>nA-</i>	Pp <i>bE-</i> for	Stative <i>nE-</i>	Perfect <i>mE-</i>
<i>permeable lexemes</i>	cold	<i>momyiy</i>	<i>no-momyiy</i>	<i>bo-momyiy</i>	<i>no-momyiy</i>	<i>mo-momyiy</i>
	night	<i>qōñ</i>	<i>nō-qōñ</i>	<i>bō-qōñ</i>	<i>nō-qōñ</i>	<i>mō-qōñ</i>
	think	<i>dēmdēm</i>	<i>nē-dēmdēm</i>	<i>bē-dēmdēm</i>	<i>nē-dēmdēm</i>	<i>mē-dēmdēm</i>
<i>blocking lexemes</i>	work	<i>mgumgu</i>	<i>na-mgumgu</i>	<i>be-mgumgu</i>	<i>ne-mgumgu</i>	<i>me-mgumgu</i>
	sleep	<i>mtimtiy</i>	<i>na-mtimtiy</i>	<i>be-mtimtiy</i>	<i>ne-mtimtiy</i>	<i>me-mtimtiy</i>
	want	<i>myōs</i>	<i>na-myōs</i>	<i>be-myōs</i>	<i>ne-myōs</i>	<i>me-myōs</i>

This means that all lexical items of Mwotlap, without exception, fall into either of two morphological categories, defined in relation with the vowel-shifting process. It is a ‘permeable’ lexeme, if it allows its first vowel to be copied onto the preceding shifting prefix: e.g. *qōñ* → *nō-qōñ*, *bō-qōñ*, *mō-qōñ*, but never \**na-qōñ*. Conversely, a lexeme will be labeled ‘blocking’ if it is incompatible with vowel shifting; in this particular case, the prefix can only take its own default vowel: e.g. *myōs* → *na-myōs*, *be-myōs*, *me-myōs*, but never \**nō-myōs*. The feature [permeable] vs. [blocking] is carried by each lexeme in the lexicon – unless it can be predicted by the theory; the latter issue will be discussed in 2.5.

At this point of the discussion, two remarks may be made. First, this rule definitely helps answer the question about the nature of the vowels in a word like *na-naw* ‘sea’ (*Table 3*). Since we are no longer dealing with irregular ‘forms’, but with irregular lexemes, then a word like *le-naw* ‘in the sea’, which has blocked the vowel-shifting process, proves the root *naw* belongs to the set of ‘blocking lexemes’; as a consequence, the *a* of the article on *na-naw* cannot be due to vowel assimilation, and is necessarily an instance of the default vowel of the prefix. The same logic applies to another ambiguous form like *le-gmel* ‘in the men’s house’: since the test with the article *nA-* gives *na-gmel* instead of \**ne-gmel*, then it becomes obvious that the root *gmel* is blocking, and that the first *e* in *le-gmel* comes from the prefix itself. Thanks to this test, all the ambiguous forms of Mwotlap can easily be solved.

Second point: only blocking lexemes can show the underlying form of each prefix; and this turns out to be essential even to morphosyntactic analysis, since they help distinguish between the noun article *nA-* and the stative prefix *nE-*. As is clear in *Table 5*, these two prefixes always have the same surface forms when they are combined to permeable roots. This point led J. Kasarhérou to a confusion, when she presented a prefix *n-* as an “actualizing prefix” (‘préfixe actualisant’) which is “compatible with all full words of this language”, regardless of their syntactic category. In fact, deeper morphological evidence reveal that there are two distinct morphemes, one of which is compatible exclusively with nouns (*nA-*), whereas the other mainly goes with adjectives and stative verbs (*nE-*). Far from challenging the noun-verb opposition, this pair of prefixes confirms that it is quite strong in Mwotlap<sup>5</sup>.

<sup>5</sup> *Table 4* suggests that *nE-* is also compatible with nouns, though this is only true of a dozen items. On the other hand, the fact that all nouns be predicative (if they bear the article *nA-*) is not a sufficient argument to say that verbs and nouns are not separated in Mwotlap; and the existence of a few ‘precategorial roots’ like the ones presented in *Table 5* (e.g. *qōñ*, *myōs*...) is not a good argument either, as long as the major part of the syntax continues to contrast nouns and verbs.

### 2.5. Is the [blocking] feature predictable?

In summary, we have demonstrated that the process for vowel shifting only occurs in one case, i.e. the combination of a permeable lexeme + a shifting prefix:

**Table 6 - Combination <permeable lexeme + shifting prefix>**

	blocking lexeme ( ${}^{\circ}myōs$ )	permeable lexeme ( $qōn$ )
unvarying prefix ( <i>ni-</i> )	– ( <i>ni-myōs</i> )	– ( <i>ni-qōn</i> )
shifting prefix ( <i>mE-</i> )	– ( <i>me-myōs</i> )	+ ( <i>mō-qōn</i> )

Now, the next question would be to find out whether this feature of the lexeme [blocking] vs. [permeable] is predictable from the form of the root, or not. This would make it clearer whether vowel shifting belongs to the domain of phonology –if it is predictable– or of morphology –if it is not. A first answer to our question is ‘No’, i.e. the form of a root is not sufficient to predict with certainty if it will shift its vowel or not. The evidence for this claim is the following (morphological) minimal pair:

- *nē-lēt* ‘pudding’: permeable lexeme /*lēt*/
- *na-lēt* ‘firewood’: blocking lexeme / ${}^{\circ}lēt$ /.

Two homophonous roots /*lēt*/ show two divergent behaviors in relation with vowel shifting: one is permeable, while the other belongs to the set of blocking lexemes. Although there are few minimal pairs like this one, it should be enough evidence to draw the following conclusion: *It is not 100% possible to predict the feature [ $\pm$  blocking] from the root itself; its value is assigned to each root within the lexicon*. In our presentation of Mwotlap lexemes, a blocking root will be preceded with a small symbol (°).

However, though there may not be a systematic rule for this prediction, we can find at least a very strong tendency in Mwotlap lexicon. It appears that permeable lexemes are generally roots *beginning with only one consonant*, whereas blocking lexemes normally *begin with two*. Table 7 illustrates this contrast between ‘CV roots’ and ‘CCV roots’:

**Table 7 - Regular correlation between root structure and vowel shifting**

permeable lexemes: CV-		blocking lexemes: CCV-	
<i>wis</i>	owl	${}^{\circ}dyē$	wait
<i>lēt</i>	pudding	${}^{\circ}twoyig$	easy
<i>siseg</i>	play	${}^{\circ}myōs$	want
<i>vēytitit</i>	fight	${}^{\circ}mtig$	coconut
<i>yō</i>	two	${}^{\circ}vnō$	country
<i>vap</i>	say	${}^{\circ}hyo$	long
<i>Motlap</i>	Mwotlap	${}^{\circ}blekat$	play cards

This table means that all CV- roots (first column) are regularly permeable, i.e. allow their first vowel to copy onto the prefix: e.g. *ni-wis* ‘owl’, *vō-yō* ‘two’, *to-Motlap* ‘from Mwotlap’; conversely, CCV- lexemes normally block the process of vowel shifting, thus being always associated with the default vowel of the prefix: e.g. *nE-* +  ${}^{\circ}twoyig$  → *ne-twoyig* ‘is easy’; *nA-* +  ${}^{\circ}vnō$  → *na-pnō* ‘country, island’.

## 2.6. A new sort of exceptions

This list could go on with thousands of lexemes: the tendency we have just mentioned is true for 100% of the verbs, adjectives and numerals, and about 95% of the nouns. In this case, it makes sense to speak about a phonological *rule*, for which there are a certain number of exceptions (< 5% of the nouns). But it must be clear that we are not following the same logic as previous authors, such as Codrington and Kasarhérou, who considered as ‘exceptions’ all roots which did not shift their vowel. These roots, which represent half of the lexicon, are now integrated in a newly-defined rule, under the name of ‘blocking roots’.

Now, what we consider to be exceptions, is a much smaller set of nouns, which do not correspond to the correlation between phonotactic structure of the root (CV- vs. CCV-) and compatibility with vowel shifting. These exceptions can be either permeable roots, which have the unexpected form CCV- (a dozen nouns); or blocking roots, which start with only one consonant (up to forty nouns). The most common of these irregular lexemes are presented in *Table 8*:

**Table 8 - A few exceptional nouns: CCV- permeable & CV- blocking roots**

permeable lexemes: CCV-		blocking lexemes: CV-	
<i>qti-</i>	head	° <i>he-</i>	name
<i>vni-</i>	skin	° <i>lo</i>	sun
<i>qñi-</i>	destiny	° <i>lēt</i>	firewood
<i>tqē</i>	garden	° <i>tō</i>	mountain
<i>ñyuñyu-</i>	snout	° <i>yen</i>	turmeric
<i>blēit</i>	plate	° <i>hōm</i>	Wrasse fish
<i>skul</i>	school	° <i>lo-</i>	inside

For instance, the word for ‘garden’ *tqē*, with the article *nA-*, does not give the expected \**na-tqē*, but an unpredictable form *nē-tqē*; in the other direction, the word for ‘mountain’ is not \**nō-tō*, as would be normal for a CV- root, but *na-tō*. Incidentally, we notice that among the morphological pair *nē-lēt* / *na-lēt* mentioned above, the first root is regular, whereas the second one belongs to the set of exceptional blocking lexemes (see *Table 8*).

The origin of most of these exceptions will be discussed in §3.3. What we would like to develop right now, is a historical hypothesis which would account for the major morphological split between permeable and blocking roots, and for its regular correlation with phonotactic structure. After this diachronic approach, we will focus on a synchronic representation of vowel shifting in Mwotlap.

## 3. A diachronic account for vowel shifting in Mwotlap

### 3.1. Historical stress and syllable loss in Mwotlap

The question is: what historical processes can account for vowel shifting in Mwotlap; and especially, why do CCV- roots systematically block this rule, while only CV- roots allow it? To answer this question, it is useful to remember that former linguistic stages of Mwotlap followed Proto Oceanic in having mainly open syllables of the form (C)V (Ross 1998: 17). The closed (CVC) syllables of Mwotlap are obviously an innovation, and were evidently formed through the loss of unstressed vowels. Basic stress rules of pre-Mwotlap

included a primary stress on the penultimate syllable of the phonological word, and secondary stresses on every second syllable towards the left of the word, e.g. *\*Mótaláva*.

In more recent times, all post-tonic syllables were deleted, causing closed (CVC) syllables to appear. Here are a few place names:

<i>*Mótaláva</i>	<i>&gt; Mōtlap</i>	'Mwotlap island'
<i>*á-Lakóna</i>	<i>&gt; Alkon</i>	'Gaua island'
<i>*á-Ravéñā</i>	<i>&gt; Ayveñ</i>	'Ravenga islet'
<i>*á-Mosína</i>	<i>&gt; Am̄sēn</i>	'Mosina village'
<i>*á-Valúwa</i>	<i>&gt; Aplōw</i>	'Valuwa village'
<i>*á-Vanúa-láva</i>	<i>&gt; Apnōlap</i>	'Vanua Lava island'

These examples show how words with four open (CV) syllables were eventually reduced to two closed (CVC) syllables, or from six to four. Other words shorten from two to one syllable (\*CVCV > CVC); other cases will be discussed below. In all these examples, only stressed syllables were retained in Mwotlap, while unstressed post-tonic vowels were definitely lost.

### 3.2. *Historical stress and vowel shifting*

Now, as far as prefixes are concerned, there were two possibilities. In a given prefixed word, either the stress would strike this prefix, or it would strike the following syllable – in which case the prefix would be in a pre-tonic position. The first case can be illustrated by a phrase in Pre-Mwotlap meaning '(the) country', *\*ná vanúa* (< POc \*panua); in Mwotlap we have:

<i>*ná vanúa</i>	<i>&gt; na-pnō</i>	'country, island, village'
------------------	--------------------	----------------------------

This example shows that when the prefix was accented, it maintained the quality of its own vowel, without any copy occurring: in modern Mwotlap, this corresponds to what we have been calling the 'default vowel' of the prefix. On the other hand, since the first vowel of the radical was in a post-tonic position, it regularly lost its vowel, thus defining a new root of the form CCV-. This hypothesis is the most powerful to explain why CCV- roots systematically block the process of vowel shifting: the genesis of such roots implied necessarily a stressed prefix, and hence no vowel assimilation at all.

The second case, i.e. the unstressed prefix, occurred whenever the phonological word (including the prefix) had an odd number of syllables, e.g.

<i>*na vánua-na</i>	<i>&gt; nē-vēna-n</i>	'his country' <sup>6</sup> ,
---------------------	-----------------------	------------------------------

In this situation, the stress on the first syllable of the root (*\*-va-*) caused it to survive in the modern form (*-vē-*)<sup>7</sup>, and this evolution is the origin of all CV- roots in Mwotlap. On the

---

<sup>6</sup> Notice that these two words are considered by speakers to be totally distinct from one another; the translation of 'his country' uses now the alienable pattern *na-pnō no-no-n*, and the word *nē-vēna-n* is a poetic word for 'his fatherland'. We have discovered up to eight etymological 'doublets' of this kind in Mwotlap (François 1999: 456), all of which are opaque to the speakers.

other hand, the unstressed (pre-tonic) vowel of the prefix, after maybe becoming a schwa, proved its articulatory weakness in assimilating totally to the next syllable. Notice that the total deletion of the prefix vowel (*\*nvēnan*) was impossible, because of the phonotactic rules applying in Mwotlap (see §1.3). In a way, vowel shifting could be considered as a special case of vowel epenthesis involving prefixes: an underlying form like *\*n-vēna-n* surfaces as *nē-vēna-n*, following *\*mtiy* → *mitiy* ‘sleep’.

Further evidence of both evolutions are presented in *Table 9*:<sup>8</sup>

**Table 9 - Former stress accounts for vowel shifting**

	<i>POc etymon</i>	<i>Pre-Mwotlap</i>	<i>Mwotlap</i>	<i>meaning</i>
<i>blocking lexemes</i>	*kasupe	*ná gasúwe	> na-ghōw	rat
	*kuRita	*ná wuríta	> na-wyēt	octopus
	*tob <sup>w</sup> a	*ná toqá-ku	> na-tqe-k	(my) bell
	* <sup>n</sup> daRaq	*ná dará-ku	> na-nye-k	(my) blood
<i>permeable lexemes</i>	*b <sup>w</sup> awe	*na qóe	> no-qo	pig
	*piRaq	*na via	> nē-vē	k.o. taro
	*suri	*na súri	> ni-hiy	bone
	*kawil	*na gáu	> ne-ge	fish-hook

To sum up, we have demonstrated that *the process of vowel shifting is historically a consequence of former<sup>9</sup> stress*: a radical would become ‘permeable’ whenever the word stress would strike its first syllable, leaving its prefix unstressed. This historical explanation accounts for the correlation, on the one hand, between permeable roots and CV-structure, and on the other hand, between blocking roots and CCV- structure in modern Mwotlap.<sup>10</sup>

### 3.3. Accounting for exceptions

Our hypothesis proves its explanatory power in accounting not only for regular processes, but also for many exceptions. Several (modern) CV- roots which unexpectedly block vowel shifting, actually used to bear stress on their very prefix, but have recently lost an extra mora. This happened when both consonants of the root were identical, causing a geminate cluster to be shortened (\*C<sub>i</sub>C<sub>i</sub>V > C<sub>i</sub>V):

<sup>7</sup> The syllable itself remains, but the quality of the vowel may change, according to the one in next syllable. Among the many possible combinations, our examples involve \*á\_u > ē (word-internally), \*á\_u > e (word-finally); \*á\_i > e; \*ú\_a > ō; \*ú\_e > ō; \*ú\_i > i; \*í\_a > ē; \*í\_o > ē.

<sup>8</sup> Notice that we spell Pre-Mwotlap and Mwotlap according to the conventions shown in §1: thus q = /k<sup>w</sup>/; d = /d/; n̄ = /ŋ/; g = /ɣ/; ē = /ɪ/; ō = /ʊ/. In Mota, a language using the same conventions, the lexical items of *Table 9* are *gasuwe*, *wirita*, *toqa-k*, *nara-k*, *qoe*, *via*, *sur*, *gau*, with no article (Codrington 1896).

<sup>9</sup> Modern Mwotlap words are systematically accented on the last syllable, not on the penultimate, contrary to what has been stated in other articles (Crowley forthc.). The evolution of stress from penultimate to final syllable is easy to understand from the deletion rules presented here; the same path was taken from Latin to French, e.g. L *cīvitāte(m)* [ki'wi'ta:te] > F *cité* [si'te] ‘city’.

<sup>10</sup> More examples and reflections have been proposed in a former paper (François 1999).

<i>*ná reréña</i>	> <i>*na-rreñ</i>	> <i>*na-reñ</i> > <i>na-yeñ</i>	‘turmeric’
<i>*ná loló-na</i>	> <i>*na-llo-n</i>	> <i>na-lo-n</i>	‘the inside of it’
<i>*ná sasá-ku</i>	> <i>*na-sse-k</i>	> <i>*na-se-k</i> > <i>na-he-k</i>	‘my name’

This happened, too, when either the first or the second syllable of the root happened to have no consonant (\*V instead of \*CV), thus bringing about a ‘pseudo’ CV root which contained an ‘invisible’ consonant. Examples of [C]CV roots include:

<i>*ná alítō</i>	> <i>na-lēt</i>	‘firewood’
<i>*ná alóa</i>	> <i>na-lo</i>	‘sun’
<i>*ná añári</i>	> <i>na-ñey</i>	‘Canarium almond’
<i>*ná awúa</i>	> <i>na-ō</i>	‘turtle’

while examples of C[C]V roots would be as follows:

<i>*ná taiúwe</i>	> <i>na-tō</i>	‘mountain’
<i>*ná saúma</i>	> <i>na-hōm</i>	‘Wrasse fish’
<i>*té Rouá</i>	> <i>te-Yō</i>	‘from Roua island’

### 3.4. *Emergence of a phonological rule in synchrony*

As would be expected, the same process happened to verb roots, bringing about blocking CV lexemes:

<i>*mé maúri</i>	> <i>*me-mir</i>	‘(it) lived, grew’ (POc <i>*maqurip</i> )
------------------	------------------	---

But remarkably, all resulting lexemes were felt to be exceptions to the regular (just emerging) correlation between phonotactics and vowel shifting: as a consequence, all verbs and adjectives underwent a morphological standardization, which created non-etymological permeable CV- lexemes:

<i>*me-mir</i>	→ <i>*mi-mir</i>	> <i>mi-miy</i>	‘(it) was in bud’
----------------	------------------	-----------------	-------------------

This process of standardization has been affecting all roots, except the most common nouns, which is not so surprising. However, the pressure of the phonological correlation here under discussion is still so strong, that even these exceptional noun roots are now beginning to conform to the norm, showing it to be still lively and productive. For instance, the ‘correct’ form *te-Yō* (see above) is sometimes heard *tō-Yō*, despite the etymology.

Parallel to this ongoing standardization, it is worth noticing that loanwords are often – though not always – forced into the same correlation between phonotactic structure of the root and compatibility with vowel shifting. For example, a CV- words like *doctor*, when combining with *nA-* article, will have its vowel copied onto the prefix: *no-dokta*; conversely, the word *policeman* gets its first vowel deleted to form a CCV- root, which will block the process of vowel shifting: *na-qlismen*.

The latter evidence suggests that the historical explanation, however powerful it may be to account for present data, must always be completed by a synchronic presentation. Present-day speakers of any language do not just make a passive use of inherited paradigms, which would only find their logic in diachrony. Instead, each generation endeavors to figure out formal similarities and constant correlations out of the amount of forms they are supposed to be using. This is how productive rules emerge, either in phonology,

morphosyntax or semantics, in such a way that the diachronic path leading to the present situation is totally ignored. Because of functional constraints in memory and ease of processing, a new consistent system is built up, slowly but surely.

Most of the time, however, this kind of standardization process is still ongoing in the language, in such a way that only a part of the lexicon –normally the less commonly used– has already been conformed to the emergent rule, whereas several items are used frequently enough to resist for a longer period. This is what happens with about fifty nouns in Mwotlap, which have not yet undergone the same overall revision that verb or adjective roots have.

#### 4. Synchronic account of vowel shifting: a multilinear approach

The aim of next paragraph is to give a synchronic description of the most productive rule for vowel shifting in modern Mwotlap.

##### 4.1. *Distinguishing tiers*

In order to achieve a formal representation of vowel shifting in Mwotlap, it makes sense to distinguish between at least two layers, one for consonants and one for vowels. Indeed, it has already been shown that Mwotlap phonology allows only vowels, not consonants, to copy, assimilate or migrate from one place to another within the phonological sequence.

This idea recalls the presentation that Nick Clements made about Kolami, a Dravidian language in which such words as *kinik*, *suulup*, *melep*, *ayak*, are supposed to follow a "rule for propagation of vowel nodes" (Clements 1991). For this purpose, Clements makes use of what a famous article by McCarthy (1989) called *planar V/C segregation*, to account for data in some Semitic and American Indian languages. Here is Clements' presentation of McCarthy's conceptions:

When [the template] is introduced derivationally, consonants link to it on one family of planes and vowels link to it on another (non-intersecting) one. At this point, consonants and vowels are entirely segregated in phonological representations, and are brought together only by the later process of tier conflation which ‘folds’ the consonant and vowel planes together.

Vowels and consonants organize into two different tiers, where they first follow their own specific rules; then both tiers eventually ‘conflate’ to conform to the *syllabic template* of the language, if there is one. Mwotlap template has already been presented in §1.3: the basic syllabic pattern of this language is |CVC|, with both Cs being optional.

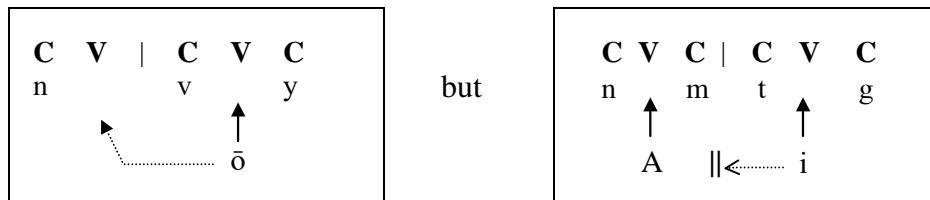
Now, as far as vowel shifting is concerned, let us sum up the rule which affects more than 95 per cent of the lexemes:

- roots beginning with one C ('CV roots') are permeable, i.e. allow their vowel to migrate to a shifting CV- prefix;
- roots beginning with two Cs ('CCV roots') are blocking roots, i.e. prevent their vowel from migrating to the prefix.

An elegant way to represent this phonological process, would be to say that a "vowel node" can propagate to the left of the word boundary, provided it has only a single C to cross

over; conversely, a cluster of two consonants behaves as a "blocking node" (Clements 1991), which hinders this vowel propagation.

Let us contrast the behavior of two nouns regarding vowel shifting: one is permeable *nA- + vōy* → *nō-vōy* 'volcano'; the other one is a blocking lexeme *nA- + °mtig* → *na-mtig* 'coconut'. In the first case, the vowel of the radical surfaces not only in its own position, but copies also to the left; in the second case, a cluster of two Cs makes the cloning process impossible:



In the remainder of this paper, our main point of interest will be the theoretical nature of the 'default vowel' of the prefix. What kind of phoneme is this, which sometimes surfaces and sometimes doesn't? The point is to escape from the feeling of having a *deus ex machina* phoneme which only exists when needed.

#### 4.2. French 'liaison' and the notion of floating phoneme

Some help may be found thanks to a typological insight on similar processes in other languages. French 'liaison' is a rule which governs surfacing vs. non-surfacing of certain word-final consonants, according to the phonotactic structure of the following word. For example, the feminine form of adjective 'small' is spelt *petite*, phonologically /pətɪ/ with a final /t/ which must always be pronounced –we shall call it a 'true' phoneme; two synonymous phrases meaning 'my girlfriend' are

- *ma petite copine* /mapəti<sub>t</sub>kɔpin/
- *ma petite amie* /mapəti<sub>t</sub>ami/.

Conversely, the masculine form *petit* shows a special /t/, which will surface only if the second word begins with a vowel: e.g. two phrases for 'my boyfriend' are

- *mon petit copain* /mõpətikopɛ/ without a /t/;
- *mon petit ami* /mõpəti<sub>t</sub>ami/ with a /t/.

The rule for liaison, which concerns the masculine form in the last example, has been described, in a multilinear framework (Encrevé 1988), as involving two fundamental notions:

- first, the notion of *syllabic template*, which consists of a (more or less constraining) string of Cs and Vs;
- second, the notion of *floating phoneme*, whose main property is to surface only on the condition that a slot has been left empty in the template, after other phonemes have taken their place.

In the case of our word *petit*, we can consider there is a *floating consonant T* at the end of the underlying masculine form /pətiT/ – vs. feminine /pətɪ/. This means that this T will

## Vowel shifting and cloning in Mwotlap

only surface if a Consonant slot has been left empty by other ‘true’ phonemes, within the syllabic template. The latter situation occurs when the next word begins with a vowel, thus leaving an empty C-slot, for the final T to slide into:

<table style="margin-left: auto; margin-right: auto;"> <tr><td>C</td><td>V</td><td>C</td><td>V</td><td>C</td><td>V</td><td>C</td><td>V</td></tr> <tr><td>p</td><td>ə</td><td>t</td><td>i</td><td>k</td><td>o</td><td>p</td><td>ɛ</td></tr> <tr><td colspan="4"></td><td style="text-align: center;">#</td><td colspan="3"></td></tr> <tr><td colspan="8" style="text-align: center;">T</td></tr> </table>	C	V	C	V	C	V	C	V	p	ə	t	i	k	o	p	ɛ					#				T								but	<table style="margin-left: auto; margin-right: auto;"> <tr><td>C</td><td>V</td><td>C</td><td>V</td><td>C</td><td>V</td><td>C</td><td>V</td></tr> <tr><td>p</td><td>ə</td><td>t</td><td>i</td><td></td><td>a</td><td>m</td><td>i</td></tr> <tr><td colspan="4"></td><td style="text-align: center;">/</td><td colspan="3"></td></tr> <tr><td colspan="8" style="text-align: center;">T</td></tr> </table>	C	V	C	V	C	V	C	V	p	ə	t	i		a	m	i					/				T							
C	V	C	V	C	V	C	V																																																											
p	ə	t	i	k	o	p	ɛ																																																											
				#																																																														
T																																																																		
C	V	C	V	C	V	C	V																																																											
p	ə	t	i		a	m	i																																																											
				/																																																														
T																																																																		

### 4.3. Floating vowels in Mwotlap

Our aim is not here to discuss French liaison, but to show how relevant Pierre Encrev  's assumptions are for Mwotlap data. The behavior of this ‘floating consonant’ in French, which sometimes surfaces and sometimes does not, makes it indeed very similar to the ‘default vowel’ of shifting prefixes in Mwotlap. The vowel /a/ of noun article, and /e/ of other prefixes (see *Table 4* in §2.3), can be described here as *floating vowels*. This means that these vowels will surface only if a V-slot is left empty in the syllabic template, once all other phonemes have been realized.

Three situations are possible:

*The V-slot has already been taken by a full vowel*

This is typically the case when the radical begins with a V, e.g. *ulsi* ‘summit’:

<table style="margin-left: auto; margin-right: auto;"> <tr><td>C</td><td>V</td><td>C</td><td> </td><td>C</td><td>V</td><td>(C)</td></tr> <tr><td>n</td><td>u</td><td>l</td><td></td><td>s</td><td>i</td><td></td></tr> <tr><td colspan="4"></td><td style="text-align: center;">#</td><td colspan="2"></td></tr> <tr><td colspan="7" style="text-align: center;">A</td></tr> </table>	C	V	C		C	V	(C)	n	u	l		s	i						#			A						
C	V	C		C	V	(C)																						
n	u	l		s	i																							
				#																								
A																												

All full-right phonemes start taking their position into the CVC|CVC pattern; then no room is left for the floating vowel. This example helps underline an essential point regarding theory: a floating phoneme may take empty slots in a given syllabic template, but it cannot create one; thus *\*na-ulsi* is excluded. This is precisely what opposes them to ‘true’ phonemes, which necessarily have a slot of their own. The following examples will help us further build on this theory.

*The floating vowel of the prefix is superseded by a vowel shifted from the root*

Another situation is when the radical is ‘permeable’, i.e. normally starts with one consonant. In this case, the first V of the radical migrates to the prefix(es), and supersedes the floating vowel.

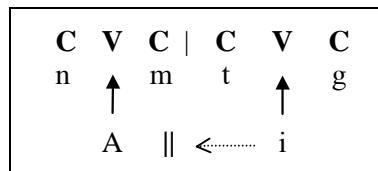
<table style="margin-left: auto; margin-right: auto;"> <tr><td>C</td><td>V</td><td>(C)</td><td> </td><td>C</td><td>V</td><td>C</td></tr> <tr><td>n</td><td></td><td></td><td></td><td>v</td><td></td><td>y</td></tr> <tr><td colspan="4"></td><td style="text-align: center;">#</td><td colspan="2"></td></tr> <tr><td colspan="7" style="text-align: center;">A</td></tr> </table>	C	V	(C)		C	V	C	n				v		y					#			A						
C	V	(C)		C	V	C																						
n				v		y																						
				#																								
A																												

What has happened here is best represented in terms of a chronology of cognitive operations:

1. consonants (here *n-v-y*) take their own slot into the syllabic template CVC|CVC, leaving V-slots empty.
2. the (first) radical vowel (here  $\delta$ ), which is a true phoneme, automatically comes into its own slot, between *v* and *y*; only one slot is now left empty, the prefix V.
3. the lexeme vowel has priority over any other V, to fill in the empty V-slot on the prefix; yet, this happens only if it is allowed by phonotactic structures, i.e. if it is not hindered by a cluster of two Cs ('blocking node').
4. if the leftward propagation of the radical vowel was blocked during stage nb. 3, then the floating vowel eventually fills the empty V-slot (see below).

#### *The slot is free*

If operation number 3 above has failed, then the vowel-slot on the prefix is still empty when the floating vowel comes in. This is the only way the 'default vowel' of each shifting prefix may surface:



The notion of floating phoneme is therefore very useful in order to account for a vowel which belongs to the lexicon, but shows intermittently. The last section of this paper will confirm how useful this notion is to describe another phonological rule of Mwotlap, namely *vowel transfer*.

## 5. Vowel transfer: floating vowels on lexemes

Without going into too much detail, it is worthwhile looking at another phonological rule of Mwotlap, which we have labeled 'vowel transfer' (see first lines of §2). This rule involves the same eight prefixes which were listed in *Table 4* (§2.3), but different lexemes, which are not concerned with vowel shifting. These lexemes can be either nouns, adjectives or verbs, but are not more than twenty in all: the process of vowel transfer is much more limited than what we have been studying so far.

### 5.1. A mobile and intermittent vowel

The principle of *vowel transfer* is that when a (shifting) prefix is added to one of these lexemes, the first vowel of the radical does not only copy to the prefix –which corresponds to ordinary vowel shifting–, but also has a rule for deletion:

*nA-* + *hinag* → \**ni-hinag* → *ni-hnag* ‘yam’

A simpler way of presenting the same process, would be to consider that the radical vowel directly migrates to the prefix, with no need for an extra deletion rule: *nA-* + *hinag* → *ni-hnag*. This presentation implies that the V of the radical has the particular property of being mobile, which is new in our system. In order to show its special status, we will write it with an uppercase letter, e.g. *hInag* ‘yam’.

A short list of the lexemes involved by the latter rule, which we may call ‘transferable roots’, include the following:

**Table 10** - Some lexemes involved in the rule for vowel transfer

<i>lIwo</i>	big	<i>mĒnay</i>	clever
<i>hInag</i>	yam	<i>vĒhog</i>	flesh
<i>sIlātēmē</i>	centipede	<i>lĒwo-</i>	tooth
<i>tIwag</i>	one	<i>dĒlo-</i>	neck

Table 10 shows that the ‘mobile vowel’ (MV) involved is always /i/ or /ē/, i.e. high and mid-high front vowels. Consequently, former accounts of Mwotlap morphology were erroneous, when they considered<sup>11</sup> a vowel like /a/ on the article (e.g. *na-gmel* ‘men’s house’) to come from the deletion of this vowel on the radical (\**gamel* < POc \**kamaliR*); actually this never happens for /a/ or /e/, which are never more than the default vowel of the prefix itself.

‘Transferable roots’ have other uncommon properties with regard to their mobile vowel. When the prefix is unvarying, e.g. *ni-* ‘3<sup>rd</sup> singular + present’, the MV just doesn’t appear:

*ni-* + *mĒnay* → *ni-mnay* ‘(he) becomes intelligent’

Conversely, this MV will surface on the right of the first consonant, when the lexeme is unprefixed, or when its prefix has the form CVC-. Table 11 gives a summary of these rules, with the root *mĒnay* ‘intelligent, clever’. The last column shows whether the ‘mobile vowel’ (MV) appears before or after the first consonant of the radical (*C<sub>1</sub>*).

**Table 11** - Rules involving the ‘mobile vowel’ of transferable roots

prefix	prefixed word	meaning	MV / <i>C<sub>1</sub></i>
<i>nA-</i>		cleverness	<b>before</b>
<i>nE-</i>	<i>nēmnay</i>	is clever	<b>before</b>
<i>mal-</i>	<i>mal-mēnay</i>	is already clever	<b>after</b>
∅	<i>mēnay</i>	clever	<b>after</b>
<i>ni-</i>	<i>nimnay</i>	gets clever	<b>no</b>

## 5.2. Floating vowels in lexemes

This property, for a phoneme, to surface or not according to its phonological environment, reminds us again of the *floating vowels* carried by our shifting prefixes (§4.3). The

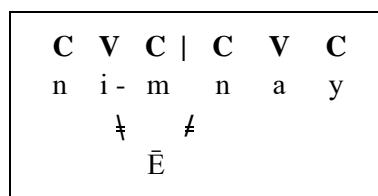
<sup>11</sup> See Codrington (1885: 311) with words like *na-bte* (Mota *patau*) ‘breadfruit’; and Crowley (forthc.) with words like *na-tmān* ‘man’.

‘mobile vowels’ of transferable roots could well be analyzed the same way, since they can be shown to share exactly the same properties. One of these, is the impossibility for floating vowels to create a slot of their own in the syllabic template: all they can do is take V-slots that have already been created by the phonological structures of the language, and which are left empty after preliminary operations have taken place.

Once again, three cases are possible:

(a) All slots have been taken by true phonemes

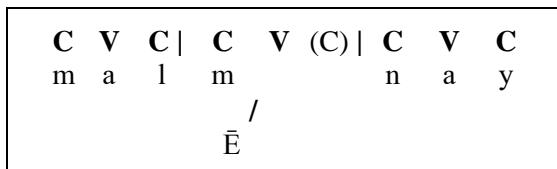
If the prefix is not a shifting prefix, but has the form CV- (e.g. *ni-* ‘Third singular present’), then both C and V, which are ‘true’ phonemes, take the first two slots of the |CVC|CVC| sequence. Afterwards, the first two consonants of the radical (*m* and *n* in our example) take their own slot, in such a way that there is no room left for the floating vowel of the lexeme to surface:



Floating vowels, by definition, cannot create their own position in the template; if they did, we would have a form like *\*ni-mēnay* instead of *ni-mnay*, and there would be no difference between this  $\bar{E}$  and a ‘true’  $\bar{e}$ .

(b) A V-slot is left empty on the right of  $C_1$

If the first |CVC| syllable has already been filled in by a CVC- prefix (e.g. *mal* ‘Complete aspect’), or if there is no prefix at all, then the first consonant of the radical ( $C_1 = m$ ) has to begin a new |CVC| syllable. Now, we know that phonotactic rules in Mwotlap exclude consonant clusters within the syllable; as a consequence,  $C_2$  (here *n*) takes the next C-slot, leaving a V-slot empty in the middle. This gives the floating vowel an opportunity to surface on the right of  $C_1$ :



(c) A V-slot is left empty to the right of  $C_1$ : competition between two floating Vs

The third case is when the lexeme takes a C- prefix (which does not exist in Mwotlap) or a shifting CV- prefix. In this case, the first three ‘true’ phonemes involved are three consonants (here *n*, *m*, *n*), which automatically fit into the syllabic template |CV-C|CVC|. The V-slot on the prefix is left empty again, allowing for a floating vowel to surface on the left of C<sub>1</sub>:

<table border="1" style="margin-left: auto; margin-right: auto;"> <tr><td>C</td><td>V</td><td>C</td><td> </td><td>C</td><td>V</td><td>C</td></tr> <tr><td>n</td><td></td><td>m</td><td></td><td>n</td><td>a</td><td>y</td></tr> <tr><td></td><td></td><td></td><td style="text-align: center;">↙ \</td><td></td><td></td><td></td></tr> <tr><td>A</td><td></td><td>Ē</td><td></td><td></td><td></td><td></td></tr> </table>	C	V	C		C	V	C	n		m		n	a	y				↙ \				A		Ē					$nA- + m\bar{E}nay \rightarrow n\bar{e}mnay$ ‘cleverness’
C	V	C		C	V	C																							
n		m		n	a	y																							
			↙ \																										
A		Ē																											

The latter example provides us with some new information. First, floating vowels in Mwotlap have the property of being mobile, which was not visible in the previous discussion; a good representation would be to say that this MV (e.g.  $\bar{E}$ ) is ‘attached’ to the first consonant of the lexeme (e.g. *m*).

Secondly, in a situation where two *floating vowels* are competing for the same empty V-slot, the vowel of the radical has priority over that of the prefix, excluding a form like *\*na-mnay*. This kind of hierarchy LEXEME > PREFIX is instructive about the way morphology is cognitively perceived.

## Conclusion

As far as the vocabulary is concerned, Mwotlap can be said to be a conservative language, since it shares many items with neighbouring Mota or with Proto Oceanic. However, historical effects of former stress have largely modified not only the shape of the words, but also the very mechanisms of the whole phonological system. Inherited structures have recently undergone several processes of standardization and reanalysis, bringing about novel phonotactic constraints –a CVC syllabic pattern– and morpho-phonemic rules, including vowel shifting. Moreover, we have demonstrated that a special sort of phoneme was created through history, namely our ‘floating vowels’; these are present in less than thirty items of the language, but their frequency makes them crucial to understand the whole morphology and grammar of the language.

Finally, we would like to put forward the idea that formal analysis is not there to confirm or invalidate a theory defined *a priori*, but must help build this theory out of the very data. Linguistic structures are already present in the way people talk, and it is neither necessary, nor scientifically satisfying, to hypothesize them out of the blue, ‘*more geometrico*’. Theory and formalism in linguistics must serve the empirical observation, and not the reverse.

## References

- ANGOUJARD, Jean-Pierre. 1997. *Théorie de la syllabe*. Sciences du Langage. Paris: CNRS.
- CLEMENTS, George N. 1991. Place of articulation in consonants and vowels: A unified theory. *Working papers of the Cornell Phonetics Laboratory* 5, 77-132. Cornell University Press.
- CODRINGTON, Reverend. 1885. *The Melanesian languages*. Oxford: Clarendon Press.
- CODRINGTON, Reverend. 1896. *Mota dictionary*. Chichester.
- CROWLEY, Terry (forthcoming). Mwotlap. In J. Lynch, M. Ross & T. Crowley (eds), *The Oceanic languages*. London: Curzon Press.

- ENCREVÉ, Pierre. 1988. *La liaison avec et sans enchaînement*. Paris: le Seuil.
- FRANÇOIS, Alexandre. 1999. Mouvements et clonages de voyelles en motlav: Entre phonologie et morphologie. *Bulletin de la Société de Linguistique XCIV*, 1, 437-486. Paris: Klincksieck.
- KASARHÉROU, Jacqueline. 1962. Les changements vocaliques de trois préfixes en motlav. *Te reo* 5, 32-34.
- MCCARTHY, John. 1989. Linear order in Phonological representation. *Linguistic Inquiry* 20.1, 71-99.
- PAWLEY, Andrew. 1993. A language which defies description by ordinary means. In W. Foley (ed.) *The role of theory in language description*. Trends in Linguistics. Berlin: Mouton de Gruyter. Pp. 87-129.
- QUICK, Phil. 2000. A non-linear analysis of Vowel harmony and Vowel harmony blocking in Pendau. In M. Klamer (ed.) *Proceedings of AFLA-7*. Amsterdam (this volume).
- ROSS, Malcolm. 1998. Proto Oceanic phonology and morphology. In M. Ross, A. Pawley & M. Osmond (eds), *The lexicon of Proto Oceanic. Vol. 1: Material culture*. Canberra: Pacific Linguistics, C-152 (pp. 15-35).

# **Chains of freedom : Constraints and creativity in the macro-verb strategies of Mwotlap**

Alexandre François

## **Abstract**

In Mwotlap (north Vanuatu), most verb phrases consist of two or more verb roots chained together  $\langle V_1-V_2\dots \rangle$ , acting like a single verb. Although a clear case of verb serialization, such phrases reveal a strong asymmetry between their free verb head ( $V_1$ ), and what appears to be little more than a post-head modifier ( $V_2$ ). Because the resulting “macro-verb” can only refer to a single action, its internal structure has to obey strict rules; this paper analyses the way the valencies of both component verbs are capable of consistently merging into that of the whole macro-verb, avoiding such things as conflicts between competing objects. Constraining though they may be, these syntactic rules turn out to be a powerful tool serving the speaker’s creativity: indeed, this paradoxical “chained freedom” brings about spectacular paths of evolution in the history of Mwotlap macro-verb strategies.

## **1. Introduction**

Quite diverse kinds of structures have been placed under the term “verb serialization”, which might well deserve different analyses. From the formal point of view, a contrast must be made between what has been acknowledged (Foley & Olson 1985) as “Nuclear-layer serialization” – of the type *I hit-die-d your brother* – and “Core-layer serialization” – of the type *I hit your brother he died*. And even after such formal precautions, it might well be also that within a single type, two different languages use similar structures to encode different semantic values – in such a way that it may not be prudent to generalize to all serializing languages the observations made for one of them. For instance, supposing nuclear-layer serializa-

tion in one language should be proved to refer to single events, it remains possible that the same syntactic device be used, in another language, to refer to several, contiguous events.

Mwotlap is an Oceanic language spoken by approximately 1,800 speakers in northern Vanuatu, Banks Is. (François 2001, 2003). More precisely, it belongs to the genetic subgroup which was labelled ‘North-Central Vanuatu’ [NCV] by Ross Clark (1985). Most of the NCV languages which have been already described have shown at least one, and sometimes two kinds of serial verb constructions: Paamese (Crowley 1987), Lewo (Early 1993), Namakir (Sperlich 1993), Ambae (Hyslop 2001) all combine nuclear-layer and core-layer serialization patterns; Araki, a language spoken in south Santo by a handful of speakers (François 2002), also has both patterns, but shows a very strong tendency for core-layer SVC. Finally, moving further northwards shows the latter structure to be less preferred than nuclear-layer SVC, in such a way that Mwotlap has virtually no example of core-layer SVC [see fn.6]. The latter facts are summarized in the following chart:

Figure 1. Different distribution of serial-verb constructions across some North-Central Vanuatu languages

some NCV languages	Core-layer SVC e.g. <i>I hit him he died</i>	Nuclear-layer SVC e.g. <i>I hit-die-d him</i>
Paama, Lewo	+	+
Araki	+	(+)
Mwotlap	-	+

The present paper will thus analyse exclusively Nuclear-layer serialization patterns in Mwotlap. In this language, it is very common – about twenty per cent of the clauses in spontaneous speech – that a single verb phrase contains not only one verb lexeme, but two or three, and up to four verb radicals, chained together within a single syntactic phrase:

- (1) *Tō kē <ni-hō mōl lok> hōw.*  
 then 3SG AO-paddle return again down  
 ‘So he paddled his way back to the west.’

After placing these structures in their syntactic context, we will analyse more specifically the way they handle such issues as valency and argument structure. This should not only allow for a better insight into the formal mechanisms of serialization, but also help formulate hypotheses on the semantic and pragmatic functions fulfilled by this strategy.

## 2. The internal syntax of VPs and Nuclear-layer serialization

It may be useful to present the basic syntax of Mwotlap clauses, so that the reader can appreciate serializing strategies in their proper context.

### 2.1. The Verb phrase

Mwotlap obeys a strict order of constituents SVO<sup>1</sup>. It is unusual within the North Vanuatu subgroup, in that neither the subject nor the object is cross-referenced on the verb form: both arguments are represented by a noun phrase or a free pronoun (sometimes a zero anaphora), with their function only indicated by their position in the clause. Mwotlap has lost the transitivizing morphology of its ancestors (e.g. POc suffix \*-i):

- (2)a. *No m-et nēk.* b. *Nēk m-et no.*  
 1SG PFT-see 2SG 2sg PFT-see 1SG  
 ‘I saw you.’ ‘You saw me.’

As far as the verb is concerned, it is obligatorily marked as finite by means of a Tense-Aspect-Mood marker. These TAM markers, of which Mwotlap has no less than twenty-five (François 2003), take the form either of a prefix, a pre-clitic, or a post-clitic. A few of these markers are discontinuous, being a combination of a prefix and a post-clitic: e.g. *et-... te* ‘realis negative’, *te-... vēh* ‘potential’; they can embrace several elements, which together form the VP. These ‘bracketing’ morphemes turn out to have a crucial role in the syntactic analysis, since they provide an efficient test for the delimitation of VPs in Mwotlap. For example, they make it easy to observe that Mwotlap VPs do not include the object:

- (3)a. *Kēy <et-galeg te> n-ēm.*  
 3PL NEG<sub>1</sub>-make NEG<sub>2</sub> ART-house  
 ‘They don’t build houses.’
- b. *Kēy <ta-galeg vēh> n-ēm.*  
 3PL POT<sub>1</sub>-make POT<sub>2</sub> ART-house  
 ‘They can build houses.’

## 2.2. Adjuncts: both a category and a function

Thanks to the same discontinuous morphemes, it is possible to characterize lexemes or phrases in Mwotlap, whether they can surface inside or outside the VP. Whereas direct objects, as well as oblique complements, are excluded from the VP, several lexemes systematically appear within this verb phrase, immediately following the verb head:

- (4) *Kōmyō* ⟨*ta-tatal tiwag lok se vēhtalōw le-mtap.*  
 2DU POT<sub>1</sub>-walk together back again POT<sub>2</sub> tomorrow in-morning  
 ‘You will be able to have a walk together once again tomorrow morning.’

Although all words in bold correspond, in English translation, to a single category called ‘adverbs’, it is obvious that they must be clearly distinguished for a language like Mwotlap: distributionally speaking, a difference must be made between those words which only fit VP-internal positions (e.g. *tiwag* ‘together’), and those which cannot enter the VP, and take the slot of oblique complements (e.g. *talōw* ‘tomorrow’). We reserve the term ‘Adverb’ for this second category, while the VP-internal words correspond to a specific function which we label ‘Adjunct’. Semantically speaking, adjuncts have the function of a head-modifier, in very much the same way as adjectives modify the nominal head of an NP.

Sentence (4) suggests that while a single verb phrase can perfectly include several adjuncts, only one verbal head is allowed at a time. To sum up, the structure of a verbal clause in Mwotlap obeys the following pattern:

Subject ⟨<sub>TAM</sub> [verbal head] + Adjuncts <sub>TAM</sub>⟩ Object + Complements

A crucial remark to be made, is that the syntactic slot of adjunct is not reserved to a few lexemes specialized in this function (“pure adjuncts”), like *tiwag* or *lok* in (4). In fact, several versatile lexemes, as well as whole categories, do fit the same verb-modifying position. For example, virtually all lexical adjectives, besides their function as noun-modifiers, can also modify semantically a verbal head, taking the slot of an adjunct:

- (5)a. *na-lqōvēn qaqa*  
 ART-woman stupid  
 ‘a foolish woman’  
 b. *Imam* ⟨*ma-hag qaqa ēwel-ēm.*  
 Dad PFT-sit stupid just in-house  
 ‘Dad is staying idly at home.’

- (6)a. *nu-quł lawlaw*  
 ART-lamp bright  
 ‘a red lamp / a shining lamp’  
 b. ⟨*Tog etet lawlaw*⟩ *kē!*  
 PROH look:DUP bright 3SG  
 ‘Stop watching her with those bright [i.e. greedy] eyes!’

Although this is less frequent, some nouns appear in the same position of adjunct:

- (7) *Kē* ⟨*ma-hag tuvusmél*⟩ *hōw.*  
 3SG PFT-sit high.chief down  
 ‘He is sitting cross-legged.’ [lit. He is sitting high-chief]  
 (8) *Tigsas kē* ⟨*et-wot vu te*⟩, *kē* ⟨*mo-wot et*⟩.  
 Jesus 3SG NEG<sub>1</sub>-born spirit NEG<sub>2</sub> 3SG PFT-born person  
 ‘Jesus Christ was not born a spirit, he was born a man.’

Through this use in the adjunct position, the set of lexical adjectives and the set of nouns both provide a stock of possible verb-modifiers, allowing new combinations to be built in order to express semantically complex processes.

## 2.3. Nuclear-layer serialization and the status of V<sub>2</sub>

It is now possible to present the serializing sentence (1) above, repeated below:

- (1) *Tō kē* ⟨*ni-hō mōl lok*⟩ *hōw.*  
 then 3SG AO-paddle return again down  
 ‘So he paddled his way back to the west.’

This sentence may be seen as a typical instance of nuclear-layer serialization: a single verb phrase includes more than one verb radical – here two. Tense-aspect-mood markers only appear once, affecting the complex verb phrase as a whole: prefixes come before the first verb, and post-clitics appear after the last verb or the last adjunct. Nothing can intervene between two serialized verbs, especially no object phrase; the object of the first verb V<sub>1</sub>, if any, is either left implicit, or becomes the object of the whole verb phrase:

- (9)a. *Tali <mi-tit tō> Kevin.*  
T. PRET<sub>1</sub>-punch PRET<sub>2</sub> K.  
‘Tali punched Kevin.’
- b. *Tali <mi-tit teñteñ tō> Kevin.*  
T. PRET<sub>1</sub>-punch cry:DUP PRET<sub>2</sub> K.  
‘Tali made Kevin cry by punching him.’

The issue of argument structure and valency will be addressed in Section 3. At this point of the discussion, it may be relevant to ask the following question: in serializing structures like (1) and (9), should we consider that the verb phrase contains two verbal heads? or is the first verb still the unique head, whereas the second one would be described as its adjunct?

The right answer to this question seems to be the second one: in a sequence  $\langle V_1+V_2 \rangle$ , the serialized verb  $V_2$ , far from being a second head, is nothing more than an adjunct to the preceding verb  $V_1$ . The first reason for this claim is the high parallelism between serialized verbs and adjuncts: they take the same syntactic slot in the verb phrase, and both have the semantic function of modifying the first verb  $V_1$ . Whatever its internal complexity, the whole VP ( $V_1+\text{adjunct}$ , or  $V_1+V_2$ ) can be understood as a mere development of  $V_1$ , with the same basic lexical meaning: for example, *hag qaqa ēwē* ‘sit just idly’, as well as *hag tuvusñel* ‘sit cross-legged’, are nothing more than two possible ways to *hag* ‘sit’; *etet lawlaw* ‘watch greedily’ refers to a way of watching; *hō mōl* ‘paddle back’ is an instance of paddling; and *tit teñteñ* ‘knock to make cry’ is a sub-type of *tit* ‘knock’, but not a sub-type of *teñ* ‘cry’.<sup>2</sup>

The latter remark can be reworded in more narrowly syntactic terms: the verbal head of a VP necessarily takes the same subject as the whole phrase of which it is the head. Thus *Tali* in (1) is both the subject of *tit* ‘knock’ and of the whole VP *mi-tit teñteñ tō* ‘knocked in such a way to make cry’. On the contrary, the following verbs in an SVC are not subject to the same syntactic constraint regarding their subject: although both verbs in (1) do have the same subject, this is not the case in (9), where the only suitable head is the first verb  $V_1$ .

Among other arguments which will not be detailed here, the asymmetry which we claim exists between  $V_1$  and  $V_2$  is confirmed by a difference in their lexical inventory. Whereas all verb lexemes of Mwotlap can be the head ( $V_1$ ) of a serializing VP, the subsequent position ( $V_2$ ) is restricted to a much smaller set of verbs, probably a few dozen; for example, such common verbs as *van* ‘go’, *lep* ‘take’, *et* ‘see’ never appear in the position of

$V_2$ . And even the verbs which can fit both positions ( $V_1$  or  $V_2$ ) sometimes show traces of asymmetry in their forms: e.g. the verb ‘know’ has the form *ēglal* when found in  $V_1$ , but *vēglal* when in  $V_2$ ; the duplicated form of the verb *sok* ‘search’ is *soksok* when in  $V_1$ , but *sosok* when in  $V_2$ , and similarly *tēy* ‘hold’ duplicates regularly as *tēytēy* when it is the head of the VP, but as *tētēy* when it is used as a head-modifier. All these remarks tend to demonstrate that what looks, at first sight, like a simple chain of verbs [ $V_1-V_2-V_3\dots$ ] placed on the same level, involves in fact an asymmetrical relationship, that of a head followed by its modifiers.

Consequently, the best way to analyse serial verbs in Mwotlap follows the model Head + Adjunct.<sup>3</sup> This pattern fits well in the category identified by Durie (1997) as asymmetrical serial constructions, whereby a single verbal head ( $V_1$ ) is modified by a limited set of verbal modifiers ( $V_2$ ).

#### 2.4. One or several actions?

It is perfectly possible that this analysis of Mwotlap does not match the structures of other serializing languages, in which SVCS would basically allow for a string of successive actions performed by the same subject. For example, the following sentence in the Papuan language Barai<sup>4</sup> could well lead to the opposite conclusion, i.e. one VP having several heads:

- BAR E *ije fu a-nafa-fu-o kan-ia buvua i.*  
man the 3SG child-PL-3SG-POSS kill-3PL cut.up eat  
‘The man killed, cut up (and) ate his children.’

Now, the reader must realize that such a string of actions<sup>5</sup> would never be coded by a serial structure in Mwotlap. Contrary to what is suggested by the general label ‘serial verbs’, this language will combine verbs in a single VP to refer to a single action, and hardly ever more. When Mwotlap needs to describe a series of actions, it does not use serialization, but coordination, by means of such conjunctions as *ba* ‘and’ or *tō* ‘so, then’ – in a way very similar to European languages.

The following example should help fix this important point. It is the narration of a series of successive actions undertaken by an individual (a healer called Boyboy) within a short period of time. Although this is typically the kind of context in which many serializing languages would make use of serial verbs, it is remarkable that Mwotlap codes all these actions by means of distinct clauses, separated by prosodic pauses<sup>6</sup> and/or coordinators (underlined): all these devices are typical of non-serializing languages.

- (10) *Bōybōy me-yem, ma-hap, mē-hēw tēy wonwon;*  
 B. PFT-climb PFT-pick PFT-descend hold intact:DUP  
 ‘B. climbed (up the coconut-tree), picked (some coconuts), brought them down intact;  
*tō kē ni-ey, tō kē ni-van tēy me l-ēm;*  
 then 3SG AO-husk then 3SG AO-go hold hither in-house  
then he husked them, then he brought them home;  
*kē ni-tot nē-tēnge nan, kē ni-van tēy me,*  
 3SG AO-chop ART-leaf ANA 3SG AO-go hold hither  
 he cut the relevant (medicinal) leaves, he brought them here;  
*tō ni-bōl madamdaw nō-gōygōyi qētēnge nan, ...*  
 then AO-hammer soft ART-roots plant ANA  
then he crushed their roots soft, ...  
*tō lep me tō nok in tō nok wē*  
 then take hither then 1SG AO:drink then 1SG AO:good  
then gave it to me, then I drank it, then I got better,  
*tō ni-bah.*  
 then AO-finish  
and this is it.’

Nevertheless, serialization patterns are not totally absent from the last citation; they appear in bold. In each case, what we observe is a single VP, inflected with a single TAM prefix (*mē-*, *ni-*), and including more than one verb radical – which corresponds well to the formal definition of ‘serialization’. But from the semantic point of view, we claim that each of these serializing VPs points to a single, minimal action, with no possibility to split it into distinct phases in time. Here they are repeated:

- (11) *Bōybōy <mē-hēw tēy wonwon>.*  
 B. PFT-descend hold intact:DUP  
 ‘Boyboy brought them down without-breaking-them.’
- (12) *kē <ni-van tēy> me l-ēm.*  
 3SG AO-go hold hither in-house  
 ‘He brought them home.’
- (13) *tō <ni-bōl madamdaw> nō-gōygōyi qētēnge nan*  
 then AO-hammer soft ART-roots plant ANA  
 ‘then he softened the roots of the plant by hammering them’

Without going into too much detail here, it is instructive to notice that the internal structure of serializing VPs in Mwotlap confirms our claim that

they cannot be interpreted as a series of actions. Indeed, it is common in most serializing languages, to translate such a verb as Eng. ‘bring’ by a string of two actions  $V_1 = \text{‘take’} + V_2 = \text{‘come’}$ ; e.g. pidgin Bislama follows iconically the order of the two phases:

- BSL *Hem i karem kokonas i kam.*  
 3SG PRD carry coconut PRD come  
 ‘He brought a coconut.’

In these languages, it is still possible to consider this string of two verbs as reflecting two successive (phases of) actions. On the contrary, Mwotlap codes the same idea using a non-iconic order of terms [ $V_1 = \text{verb of movement} + V_2 = \text{‘hold’}$ ], in such a way that the “*n* verbs, *n* actions” interpretation becomes impossible. The only reading possible for sentences (11) and (12) corresponds not to successive actions, but to simultaneous facets of a single action: Boyboy descends (from the tree) holding (the coconuts) and keeping them in one piece.

## 2.5. Summary: An optical illusion

We can now summarize the results of these first observations about Mwotlap. In this language, a single verb phrase may include more than one verbal lexeme at a time, with no other element intervening. The surface pattern  $\langle V_1+V_2+V_3\dots \rangle_{VP}$  recalls similar strings of verbs in certain languages – like Tariana (Aikhenvald 1999) – and suggests the term of ‘(nuclear-layer) verb serialization’ for Mwotlap.

However, a deeper analysis shows that the term ‘serial verb’ may well be an optical illusion. First, there is a formal and semantic asymmetry between  $V_1$ , the unique head of the verb phrase, and the following verbs, whose basic role is to modify this head. In this sense, serialized verbs enter a syntactic slot we have called ‘adjunct’; far from being exclusive to verbs, the position of adjunct is also open to adjectives and nouns, plus many lexemes (“pure adjuncts”) exclusive to this function. From the semantic point of view, a serial verb string in Mwotlap cannot refer to several distinct actions – in which case, coordination is used – but to a single action, undertaken by one subject at a given point in time. The internal complexity of these SVCs allows coding this single action under several of its facets. Focusing on the issue of valency and argument structure, the second section of this paper will examine the way this complexity is handled by the so-called ‘serializing’ strategy.

### 3. Serial verbs and transitivity issues

After this first exploration of Mwotlap SVCS, we would like to focus our attention on the relations existing between the argument structure of a serialized VP and that of its components. Indeed, we have already seen that a multi-verb VP behaves externally like any simple verb, having one subject on its left, and no more than one object on its right. Knowing that Mwotlap does not allow for double-object structures (Eng. *I gave him a pen*), nor is it possible to insert any object NP between two serialized verbs, syntactic conflicts may arise in the combination of two transitive verbs (see [o] below).

In reality, cases of valency conflict seldom occur in Mwotlap SVCS. Most of the time, what is observed is a regular capacity to merge the argument structures of two verbs into that of a new, composite ‘macro-verb’. The following paragraphs will try and establish a syntactic classification of Mwotlap SVCS, according to the argument structure of their input elements, and of the output verb phrase.

Basically, the main relevant opposition is that between intransitive and transitive verbs: for instance, we will see in which cases the combination of two intransitive verbs leads to the formation of an intransitive or a transitive VP. But for this analysis, finer criteria will be needed, such as the sameness or difference between, say, the object of  $V_1$  and the subject of  $V_2$ , etc. As a consequence, each type in the following classification will be presented with a simple formula, using small letters ( $x, y, z$ ) for arguments, and an SVO convention; e.g.  $[x-V_1]$  means ‘ $V_1$  is an intransitive verb having a subject  $x$ ';  $[x-V_2-y]$  means ‘ $V_2$  is a transitive verb having a subject  $x$  and an object  $y$ ’.

A short note is necessary here about the category “adjective”. In Mwotlap as in many Austronesian languages, adjectives follow the same syntactic patterns as intransitive verbs, in most contexts: both categories are directly predicative, they combine with the same aspect-mood markers, and so on; the only position where they can be distributionally contrasted is inside the noun phrase, since only an adjective can modify a noun directly. Because our study is concerned with serial constructions in predicative phrases, where adjectives and verbs are merged, it will here be legitimate to regard adjectives as a sub-class of intransitive verbs, following the pattern  $[x-V_1]$ .<sup>7</sup>

#### 3.1. Both verbs are intransitive

[a] 
$$\boxed{x-V_1 + x-V_2 \rightarrow x-[VP]}$$

Both verbs can be intransitive. If their individual subject is semantically the same, then the output is normally an intransitive macro-verb. Semantically, the subject can be said to perform both ‘actions’ – or, to be more precise, both facets of the same action – at the same time.

- (14) *Nok* ⟨*tag*      *mitiy*    *tusu1SG   AO:bend.down sleep   a bit  
'Let me have a nap.'*

- (15) *Gēn*   ⟨*yow*    *tig*⟩   *hōw*   *lē-bē*!  
1IN:PL   AO:jump stand down in-water  
'Let's jump into the river!'

- (16) *Inti*   ⟨*ma-kal*   *qeñen*⟩.  
son:2SG PFT-crawl disappear  
'Your baby's crawled away.'

As was mentioned before, nothing prevents us from adding to the present list those cases in which the adjunct  $V_2$  is an adjective (or “adjectival verb”), with the same semantic subject as  $V_1$ :

- (17) *Kē*   ⟨*me-teñ*   *magaysēn*⟩.  
3SG   PFT-cry   sad  
'He was crying miserably.'

[b] 
$$\boxed{x-V_1 + x-V_2 \rightarrow x-[VP]-x}$$

A rather strange pattern provides an exception to [a] above, since it shows two intransitive verbs with the same subject, resulting in a formally transitive macro-verb. The object of this VP has the same reference as its subject, which corresponds, incidentally, to the coding of reflexive verbs in Mwotlap.

In fact, this pattern [b] occurs only in familiar speech, with basically two verbs in adjunct position: *mat* ‘die’ and its slangish counterpart *mem* ‘piss’. The basic idea is that the subject  $x$  is performing an (intransitive) action  $V_1$  in such an intense manner, that it makes him metaphorically die... or, less seriously, urinate. Practically, this structure is used as a jocular intensifier for certain intransitive actions:

- (18) *Nēk* ⟨*mē-yēyē mat*⟩ *nēk aē ēgēn!*  
 2SG PFT-laugh die 2SG ANA now  
 ‘You’re laughing yourself to death!’

- (18)' *Nēk* ⟨*mē-yēyē mem*⟩ *nēk aē ēgēn!*  
 2SG PFT-laugh piss 2SG ANA now  
 (slang) ‘This makes you piss with laughter!’

In order to be understood, this isolated pattern [b] should be compared to others, more productive. For instance, it can be interpreted with reference to [e] below – except that in this case  $x = y$ . Or it may also be compared with the causative structure ([j] below), with which both verbs *mat* and *mem* are usually associated, in phrases like { $x V_1 mat y$ } ‘ $x$  kills  $y$  (through the action  $V_1$ )’ or { $x V_1 mem y$ } ‘ $x$  makes  $y$  piss (through a violent action  $V_1$ , e.g. knock down or scare)’. What is particular in (18)–(18)’, is that  $V_1$  is intransitive, and the patient coincides with the actor<sup>8</sup>.

- [c] 
$$\boxed{x\text{-}V_1 + =V_2 \rightarrow x\text{-}[VP]}$$

It sometimes happens that the logical subject of  $V_2$  is not just  $x$  (the subject of  $V_1$ ), but rather corresponds to the predicate structure  $x\text{-}V_1$  as a whole. The verb  $V_2$  comments on the manner in which the action ( $x\text{-}V_1$ ) is carried out.<sup>9</sup> This is often the case when the serialized element is an adjective, which may then be said to work as an “adverb”:

- (19) *Na-day nono-n* ⟨*me-plag lililwo*⟩.  
 ART-blood of-3SG PFT-run big:DUP  
 ‘His blood flowed abundantly.’

- (20) *No-ko e kē* ⟨*ma-mat hiywē*⟩.  
 art-pig ANA 3sg PFT-die be.true  
 ‘The pig was well and truly dead.’

In example (20), it is clear that the subject of  $V_2$  ‘be true’ is not the pig itself, but ‘the dying of the pig’: the logical structure of these sentences is thus { $V_2(V_1(x))$ }, involving a second-order predicate. See also [h].

- [d] 
$$\boxed{x\text{-}V_1 + \emptyset\text{-}V_2 \rightarrow x\text{-}[VP]}$$

Sometimes, an intransitive  $V_1$  is combined to an impersonal verb  $V_2$ , with zero-valency<sup>10</sup>, e.g. *qōñ* ‘be night’, *myen* ‘be daylight’. The latter does not affect the valency of the main verb, and adds only the meaning ‘(do  $V_1$ ) until it is night [resp. day]’.

- (21) *Tita* ⟨*ni-hag qōñ*⟩ *l-ēm.*  
 Mum AO-sit be night in-house  
 ‘Mum will stay at home all day long.’

- (22) *Kimi* ⟨*ma-lak meyen tō*⟩?  
 2PL PRET<sub>1</sub>-dance be day PRET<sub>2</sub>  
 ‘Did you dance all night long?’

- [e] 
$$\boxed{x\text{-}V_1 + y\text{-}V_2 \rightarrow x\text{-}[VP]\text{-}y}$$

Two intransitive verbs may merge into a transitive macro-verb: this happens when the semantic subject of  $V_2$  is different from that of  $V_1$ . In this pattern, the general meaning is “ $x$  performs/undergoes an intransitive action  $V_1$ , which results in another element  $y$  undergoing in turn a transformation ( $V_2$ )”; the syntactic output of this combination is a transitive macro-verb ⟨ $V_1$ - $V_2$ ⟩ with a causative meaning. Notice that the present pattern normally does not concern animate actors, but rather natural forces: with an animate subject, the feature [control] would normally result in the choice of a transitive  $V_1$ .

- (23) *Na-lo* ⟨*ni-hey simsim*⟩ *n-aes.*  
 ART-sun AO-shine melt:DUP ART-ice  
 ‘The sun melts the ice (by shining).’

- (24) *Ne-leñ* ⟨*mi-yip hal-yak*⟩ *na-kat.*  
 ART-wind PFT-blow fly-away ART-cards  
 ‘The wind blew the cards away.’

- (25) *Ni-yiy* ⟨*mi-yiy sisigoy*⟩ *na-mtig.*  
 ART-quake PFT-quake fall:DUP ART-coconut  
 ‘The earthquake made the coconut trees fall down.’

Even when it is animate, the subject is generally not agentive – but note ex. (27):

- (26) *Nēk* ⟨*mi-tig mēlēmlēg*⟩ *na-lo den kemem.*  
 2SG PFT-stand black ART-sun from 1EX:PL  
 ‘Standing as you are, you’re hiding the sun from us.’  
 [lit. You’re standing dark the sun from us.]

- (27) ⟨*Gengen maymay*⟩ *na-taybē!*  
 AO:eat:DUP strong ART-body:2SG  
 ‘Eat well, to strengthen your body.’ [lit. Eat strong your body.]

To our knowledge, pattern [e] was never witnessed in any other serializing language which has been described to date. It must be distinguished from the so-called ‘switch-subject serialization’ ([j]), since the first verb is intransitive: we would call it “low agency causative serialization”. Most remarkably, this structure contradicts the claim usually made about serialization, that “serial verbs share at least one (...) argument” (Durie 1997: 291): Mwotlap proves that two predicates, having no argument in common, can perfectly merge into a single serial verb construction.<sup>11</sup> The structure, however, is rare: our corpus shows little more than the few examples cited here.

[f] Combination of several verbs, none being transitive

Any combination of the above formulas, involving more than two verbs, leads to the expected result. We will give here only one example, which combines [c] and [d] above: the combination of an intransitive verb + an adjunct commenting on this first action + an impersonal adjunct, results in an intransitive macro-verb.

- (28) *Kōyō* <S-S.P.R. *qaqa qōñ*.  
3DU AO:roam:DUP stupid be night  
'They spend the whole day aimlessly wandering.'<sup>12</sup>

The formula corresponding to this example would be:

$$x\text{-V}_1 + =V_2 + \emptyset\text{-V}_3 \rightarrow x\text{-[VP]}$$

Such combinations are very common in everyday speech, and there even seems to be a preference for this kind of multi-verb serialization in colloquial discourse and slang.

### 3.2. Only one verb is transitive

When only one verb is transitive, then the result of the combination is invariably a transitive VP.

$$[g] \quad x\text{-V}_1\text{-}y + x\text{-V}_2 \rightarrow x\text{-[VP]\text{-}y}$$

Despite its simplicity, this pattern seldom occurs in Mwotlap: usually, the combination of a transitive head with an intransitive adjunct is interpreted as a ‘switch-subject serialization’ ([j] below), i.e. a causative structure in

which the subject of V<sub>2</sub> is understood to be *y*, not *x*. However, the following example is compatible with the interpretation [g]:

- (29) *Yē* <*ti-tiok magaysēn*> *nēk?*  
who FUT-see.off sad 2SG  
'Who will have the sad role to see you off?' [x is 'sad']  
(or maybe: 'who will see you off, you poor fellow...') [y is 'sad']

$$[h] \quad x\text{-V}_1\text{-}y + =V_2 \rightarrow x\text{-[VP]\text{-}y}$$

An intransitive verb, or often an adjective, can describe the manner of a transitive action V<sub>1</sub>. In comparison with other patterns, what appears here is that the logical argument of predicate V<sub>2</sub> is neither *x* nor *y*, but the whole event  $\langle x\text{-V}_1\text{-}y \rangle$ : this is another case of so-called ‘ambient serialization’ (see [c]).

- (30) *Na-bago* <*mi-ñit maymay*> *kē*.  
ART-shark PFT-bite hard 3SG  
'The shark bit him viciously.'
- (31) *Kēy* <*sok walēg*> *kēy le-pnō*.  
3PL AO:seek round 3PL in-island  
'They looked for them around the island.'
- (32) *Ige tamān kēy* <*tit-vasem soloteg vēste*> *na-halgay en.*  
PL man 3PL POT<sub>1</sub>:NEG<sub>1</sub>-reveal random POT<sub>2</sub>:NEG<sub>2</sub> ART-secret ANA  
'Men must not carelessly reveal the secrets (of initiation).'

For example, in (30), what is ‘hard’ is neither the shark itself nor its victim, but the whole state of affairs ‘[the way] the shark bit him’. This corresponds exactly to the logical description of adverbial modifiers in a language like English, which take a second-order predicate as their logical subject (Dik 1989: 193).

$$[i] \quad x\text{-V}_1\text{-}y + \emptyset\text{-V}_2 \rightarrow x\text{-[VP]\text{-}y}$$

In a way parallel to [d] above, it is perfectly possible to combine a transitive verb with an impersonal (“weather”) verb, leading to a transitive verb phrase. V<sub>2</sub> acts as an adverb –more precisely an “intransitive adjunct”—without affecting the valency of V<sub>1</sub>:

- (33) *Kem* <*soksok meyen tō*> *nēk.*  
 1EX:PL seek:DUP be day PRÉT 2SG  
 ‘We’ve been looking for you all night.’

[j] *x-V<sub>1</sub>-y + y-V<sub>2</sub> → x-[VP]-y*

One of the most frequent serializing patterns in Mwotlap has a causative meaning: an actor *x* acts (*V<sub>1</sub>*) upon a patient *y* in such a way that *y* undergoes the intransitive process *V<sub>2</sub>*. This structure was called ‘switch-subject serial verbs’ or ‘serial causative verbs’ by Crowley (1987:39). Sentences (9) and (29) above already illustrated this case.

- (34) *No* <*ma-kay metewot*> *na-tmān vōyō*.  
 1SG PFT-shoot injured ART-man two  
 ‘I wounded two men (by shooting at them).’

- (35) *Kōyō* <*mō-bōw liwo*> *kē*.  
 3DU PFT-bring.up big 3SG  
 ‘They brought him up (till he was big).’

- (36) *Kē* <*ni-vatne lolmeyen*> *gēn.*  
 3SG AO-teach wise 1IN:PL  
 ‘He makes us wise (through his teaching).’

The usual causative structure<sup>13</sup>, which uses *ak* ‘make’ + *V<sub>2</sub>*, can also be described as a serial verb structure:

- (37) <*Ak tog-yon*> *kē!*  
 AO:make stay-quiet 3SG  
 ‘Make him be quiet!’

But it must be noticed that Mwotlap speakers, whenever they want to express such a causative action performed by *x* upon *y*, will always prefer to use a more specific verb than *ak* ‘make’ in the first position. For example, the action of ‘softening’ will seldom be expressed *ak madamdaw* ‘make soft’: despite being correct, such a phrase is felt to be incomplete or childish. The most idiomatic strategy is to use the ‘switch-subject serialization’, by specifying the action *V<sub>1</sub>* which is supposed to result in the patient becoming ‘soft’: by hammering it, by chewing it, by pulling on it, etc. In a way, this strategy just consists in “replacing” the all-purpose verb *ak* by a semantically more specific verb, e.g. *bōl* ‘hammer’, etc. This is the case in (13), and other similar sentences:

- (38) *Kē* <*ni-kuy madamdaw*> *na-ga.*  
 3SG AO-chew soft ART-kava  
 ‘He softened the kava by chewing it.’

- (39) *Lep* <*ne-vet wiyyiyiy, ba wiyyiy madamdaw*>. *ba* <*wiyyiy madamdaw*>.  
 AO:take ART-stone grind:DUP and AO:grind soft  
 ‘You take the grinding stone, and soften (the kava) by grinding it.’

Through these examples, the serializing strategy clearly appears as a way to gather the semes included in several individual verbs, and merge them into a single macro-verb. The resulting VP is an attempt to encode the complexity of a single action by expressing several of its facets. As we have already pointed out, it would be misleading to consider these chains *V<sub>1</sub>-V<sub>2</sub>* to reflect two successive events in time (e.g. *he hammered the roots*, and then *they became soft*); rather, the two radicals form a narrow bundle, evoking the specific action of softening something by hammering it.

[k] *x-V<sub>1</sub>-y + z-V<sub>2</sub> → x-[VP]-z*

A less frequent variant of the causative pattern we have just presented also combines a first transitive verb *V<sub>1</sub>* with an intransitive verb *V<sub>2</sub>*; but the subject of *V<sub>2</sub>* is by itself a new element (*z*), corresponding neither to the subject (*x*) nor to the object (*y*) of *V<sub>1</sub>*. Since there is only one object slot available for two distinct applicants (*y / z*), we are facing here the first case of syntactic conflict between arguments.

The solution adopted by Mwotlap is normally to drop the object *y* of *V<sub>1</sub>*, thus leading to a transitive verb oriented towards *z*, the experiencer of *V<sub>2</sub>*: {*x-[VP]-z*}. Often, *y* either features as a topic in the same sentence [see also ex. (48) below], or is easily reconstructed from the discourse context:

- (40) *Tita nonon mi-gil nē-qyō-n,* <*mi-gil wawah*> *na-taqmē-n.*  
 mother his PFT-dig ART-grave-3SG PFT-dig clean ART-body-  
 3SG  
 ‘His mother dug his grave, cleaning (the earth) off his body.’  
 [lit. she dug his grave, dug clean his body]

- (41) *Kē ma-yah nē-sēm nen etō <ni-yah ōlōl>*  
 3SG PFT-scrape ART-cowrie that then AO-scrape call:DUP  
*nē-yēdēp ēgēn.*  
 ART-Pritchardia now

'And as she was scraping those cowries (to make shell-money), her scraping made the palm leaf screech.' [lit. she scraped screech the leaf]

- (42) *Kē <ma-vap lōlmeyen> nēk.*  
 3SG PFT-say aware 2SG  
 'She said (it) to let you know.' [lit. she said aware you]

The last sentence can be compared with (36) above: whereas the implicit object of *vatne* 'teach (s.o.)' was the same as the subject of *lōlmeyen*, we have here a verb  $V_1 = vap$  'say (s.th.)', whose object is different from that of the whole macro-verb: it has to be sought in the preceding context.

The combination *vap lōlmeyen* is the usual way to translate the notion 'inform s.o. of s.th.', 'let s.o. know s.th.'. This makes it easy to understand why, in a language like Mwotlap having only one object-slot, conflicts may arise as to which of the two semantic objects will be retained as such in the final structure – whether the 'news' itself (underlying object of *vap* = *y*) or the 'person informed' (underlying subject of *lōlmeyen* = *z*); (42) shows that the second choice is the right one, the object of  $V_1$  being left implicit.

Interestingly, these serializing strategies may be viewed as a powerful way to increase the density of information within the narrow limits of a single VP, hence increasing semantic explicitness – but it could also be seen, conversely, as a laconic shortcut in which much of the information has to be left unmentioned, with only a small selection of semes coming to light. The right conclusion will depend on the point of comparison: as opposed to a single verb *V*, a combination of two or more will certainly bring about new elements and richer information. But if a complex verb  $\langle V_1-V_2 \rangle$  were to be compared with a linkage of two corresponding full clauses with their own arguments, then what would appear would certainly be some kind of semantic loss and blurring. This is precisely the case with (42). '*She said aware you*', which contains more than just *She made you aware* – but less than *She said these words to make you aware*. With such serial structures, languages seem to be testing the limits between semantic explicitness and syntactic concision. This risk of uninterpretability is certainly the reason why [k] occurs seldom in Mwotlap, while multi-clause strategies are preferred.<sup>14</sup>

- [I]  $xV_1 + x\text{-}V_2y \rightarrow x[\text{VP}]\text{-}y$

This new pattern is as important as the causative one [j], but although they both lead to the same surface result, the mechanism is quite different. Often, we have a first intransitive verb, referring to a single-argument action ( $V_1$ ); but for its subject *x*, this first action can also be presented as a way to act upon a patient *y*, or as a process ( $V_2$ ) involving, in one way or another, a binary relationship between two elements *x* and *y*. The result of this is always a transitive verb phrase.

- (43) a. *Kōmyō <ta-kaka vēh>*.  
 2DU POT<sub>1</sub>-chat POT<sub>2</sub>  
 'You may discuss.' [intr.]
- b. *Kōmyō <ta-kaka gatay vēh> no.*  
 2DU POT<sub>1</sub>-chat mention POT<sub>2</sub> 1SG  
 'You may discuss about me.' [discuss mentioning me]
- c. *Kōmyō <ta-kaka tataq vēh> na-myōs.*  
 2DU POT<sub>1</sub>-chat follow POT<sub>2</sub> ART-desire  
 'You may discuss freely.' [discuss following your desire]

In these sentences, the act of 'chatting' ( $V_1$ ) is presented as involving a secondary semantic relationship between the subject *x* and another contextual element *y*. In each case, the function of the serialized verb  $V_2$  is to introduce a relational seme  $f(x,y)$  involved in the process. Thanks to the serial strategy, the speaker is able to mingle unary semes (e.g. 'x chats') and binary semes (e.g. 'x mentions y') in order to construe a satisfying representation of a given situation. To avoid syntactic conflicts between objects, only one such combination can occur in a given clause: therefore, it is not possible to say in Mwotlap *You may talk freely about me*; the speaker would have to use more than one clause to encode all these relations.

By adding a relational seme<sup>15</sup> to the main verb, the adjunct serialized here (*gatay*, *tataq*) also has the remarkable effect of increasing its valency. This has important, synchronic and diachronic, consequences, which may lead  $V_2$  to grammaticalize as a transitivity morpheme – an issue which will be detailed in the last section of this paper (4.3).

The "argument-adding" pattern under discussion is particularly developed when the head verb expresses movement [see ex. (11)-(12)], like *van* 'go', *hēw* 'go down', *vēykal* 'go up', *hayveg* 'go in', *kalō* 'go out', *hō* 'paddle'...: all these motion verbs, and many more, are attested combining with a transitive adjunct ( $V_2$ ), in order to gain one extra slot in their argu-

ment structure – that is, gain an object. The proper meaning of V<sub>2</sub> involves a relational seme between *x* and *y*:

- (44) *No <ma-van tēy> na-gasel mino.*  
1SG PFT-go hold ART-knife my  
'I took my knife away (with me).'
- (45) *No <ma-van veteg> na-gasel mino.*  
1SG PFT-go (leave) ART-knife my  
'I left my knife behind (when going).'
- (46) *No <ma-van goy> na-gasel mino.*  
1SG PFT-go (over) ART-knife my  
'I went to fetch my knife.'

In each sentence, the agent's movement (*van* 'go') involves a certain semantic relationship regarding the knife: he has it with him / he leaves it behind him / he is looking for it, etc. Note that only (44) can be said to involve a true serial structure, because *tēy* is still productive as a head verb in modern Mwotlap; as for *veteg* in (45), it used to be a serialized verb, but is now becoming a pure adjunct (4.3.2) and the adjunct *goy* in (46) was probably never a verb at all in this language. Consequently, if we are to provide a consistent analysis for such sentences as (44) to (46), the relevant syntactic category should be broadened to (transitive) adjuncts, of which (transitive) verbs are just a subset.

[m] Combination of several verbs, one of which is transitive

Finally, the algebraic formulas we have been listing can generally combine together (see [f]). This is basically done, one could argue, by bracketing pairs of verbs, proceeding from left to right – i.e. starting from the head. For example, when we find a string of three or four verb radicals V<sub>1</sub>-V<sub>2</sub>-V<sub>3</sub>-V<sub>4</sub>, it is possible to calculate the overall valency of the resulting macro-verb by considering that the rightmost adjunct has been added to a (smaller) macro-verb, having itself its own argument structure:

$$\{[(V_1-V_2) - V_3] - V_4\}$$

For instance, a famous love song in Mwotlap has the following string of head + adjuncts:

- (47) *<Lak tēy yoyon ēwē> no.*  
AO:dance hold quiet good 1SG  
'Just dance with me calmly.'

Knowing the proper valency of each constituent verb, it is easy to see that their combination regularly follows the rules for each pair of verbs:

- rule [l] x-dance + x-hold-y  
→ *x-[dance with]-y*
- rule [h] x-[dance with]-y + = quiet  
→ *x-[dance calmly with]-y*
- rule [h] x-[dance calmly with]-y + = good  
→ *x-[just dance calmly with]-y*

The result of this combination of rules, as expected, is a transitive macro-verb, whose subject is the subject of the head (by essential property of head), and whose object is the object of the only transitive verb of the string, namely V<sub>2</sub> 'hold'.

A similar combination involves the rare pattern [k] above, in which three semantic arguments had to share only two syntactic slots. Once again, the rejected element is *y* (the object of V<sub>1</sub>), which only appears as a sentence-initial topic:

- (48) *Yebek en, kemem <ōl tog-yoñ magaysēn> ige susu.*  
devil ANA 1EX:PL AO:call stay-quiet sad PL children  
[lit. That Yebek (legendary monster), we call silent sad the kids.]  
'That Yebek, we call (him) to scare the kids silent.'

This sentence can be read as a development of (37) above: instead of the all-purpose verb *ak* 'make', a more specific verb is used to express the action which is performed to achieve the goal *tog-yoñ* '(the children) be quiet'. But contrary to ordinary causative sentences [e.g. (13)], the object of V<sub>1</sub> is here different from the global undergoer (the "causee"): *parents* are naming the *monster* in order to calm the *children* down. In other words:

- rule [k] x-call-y + z-be.quiet → *x-[shut up]-z*

Now, notice the third predicate in the same series. Evidently, once a complex transitive verb has been construed, it is possible to add an extra verb V<sub>3</sub>, and thus build a new transitive macro-verb, according to rule [j] above:

- rule [j] x-[shut up]-z + z-sad → *x-[terrorize]-z*

The reader will appreciate how much concision is made possible by the use of a serial strategy: this is how a single VP can combine such unrelated notions as 'call a name', 'be silent', and 'be sad'. The sentence itself illus-

trates a common case of adult cruelty towards their children... but this is another issue.

### 3.3. Both verbs are transitive

We shall now mention the last combination patterns, those which involve two transitive verbs.

[n]  $x\text{-V}_1\text{-}y + x\text{-V}_2\text{-}y \rightarrow x\text{-[VP]-}y$

No conflict arises if both verbs share the same subject and the same object: the result is a similar macro-verb. The function of V<sub>2</sub> is obviously not a syntactic one – e.g. valency-increasing– but consists in adding some semantic features to V<sub>1</sub>.

- (49) *Kēy <et- et vēglal te> no.*  
 3PL NEG<sub>1</sub>- see know NEG<sub>2</sub> 1SG  
 ‘They did not recognize me.’

- (50) *Nok <tivig veteg bah> kē en.*  
 1SG bury (leave) PRIOR 3SG ANA  
 ‘Let me first bury (and leave) him.’

[o]  $x\text{-V}_1\text{-}y + x\text{-V}_2\text{-}z \rightarrow x\text{-[VP]-}z$

The last case we will be facing does involve a typical conflict between arguments, in a way similar to pattern [k] above. Each transitive verb has its own object, which results in two items applying for the single object-slot of the macro-verb. Once again, the usual strategy used by Mwotlap is to select the object of V<sub>2</sub> (z) to that position, thus leaving implicit the object of V<sub>1</sub> (y).

In (51) below, ‘beer’ is the semantic object of V<sub>1</sub> ‘drink’. But because V<sub>2</sub> ‘accompany’ also has its own object (z = ‘kava’), the word ‘beer’ is formally excluded from the clause, and can only be retrieved from the immediately preceding context:

- (51) *Nēk ta-kalē vēh mi ni-bia... a so <in biyiñ ēwē> na-ga.*  
 2SG POT<sub>1</sub>-chock POT<sub>2</sub> with ART-beer that is drink accompany just ART-kava  
 ‘[when drinking kava] one may “force it down” with beer: that means nothing but <drink together with> kava.’

Such a syntactic constraint, which makes it necessary for the speaker to introduce an item (here y = ‘beer’) in a preceding sentence, shows that Mwotlap Nuclear-layer SVCs must be analysed in a very different way from Core-layer SVCs, even in closely-related languages. For example, Paameese has the possibility, thanks to its Core-layer SVCs, to express each verb with its own object, in a structure which Crowley (1987) labels ‘multiple-object serialization’. Obviously it does not have the same consequences for syntax and discourse strategies. Compare Mwotlap (51) with this Paameese sentence (1987: 39):

- PAA *Inau na-mun sīn dal oai.*  
 1SG 1SG:REAL-drink gin 3SG:REAL:accompany water  
 ‘I drank gin with water.’ [lit. I drank gin it accompanied water]

We do not think the “rules” we are defining for Mwotlap have to be conceived in terms of a purely formal constraint, e.g. Mwotlap forbidding (vs Paameese allowing) the insertion of an object inside SVCs. It seems more convincing to consider that these two languages, despite their vicinity, employ two drastically different structures, including in semantic terms. In Paameese, the verb *mun* ‘drink’ remains perfectly oriented towards its own, expected object, without being much affected by the presence or absence of a serial structure – in a way typical of Core-layer SVCs. Conversely, Mwotlap SVCs do considerably affect the diathetic orientation of its verbs, in such a way that *in* ‘drink’ in (51) could be said to be no longer oriented towards its “semantic object” (‘beer’), but rather has become part of a macro-verb, the only function of which is to specify it semantically. Considering *in biyiñ* as a whole, we must admit that the serializing operation has blurred the syntactic link between ‘drink’ and ‘beer’, in such a way that the macro-verb <drink-accompany> is now semantically oriented towards the ‘kava’: it describes a certain action that the actor performs in relation with kava, e.g. increasing its effect, improving its taste, etc. In the framework of this new action, the item ‘beer’ is neither an object nor –we claim – a patient, but has now a sort of semi-presence, in the same way as a peripheral argument (instrument, locative...) could have.

The hypothesis we are suggesting, and which has theoretical consequences, is that when an SVC pattern has the effect of demoting an object from its position, then this demotion does not only take place at a syntactic ('surface-structure') level, which would leave intact its semantic role as a patient. In Mwotlap SVCS, any change altering the syntactic status of an object also makes it necessary to modify its semantic properties, in such a way that what was once a patient is redefined as a peripheral case role (e.g. instrument). Thus compare this sentence, in which *na-hat* is object and patient:

- (52) a. ⟨*Hey*⟩   *na-hat*   *anen.*  
          AO:wear   ART-hat   that  
          'Put that hat on.'

...with the following one, in which the same action (*hey* 'put on, wear') has been integrated as the first element of a macro-verb *hey goy*, and has therefore lost its object:

- b. ⟨*Hey*   *goy*⟩   *ni-qtí*   *MI*   *na-hat*   *anen.*  
          AO:wear (cover)   ART-head:2SG   with   ART-hat   that  
          'Cover your head with that hat.'

Since it has been rejected from the position of object of  $V_1$ , the item 'hat' is no longer a patient, but has been redefined as an instrument (preposition *mi*) within a new action: *lit.* 'You ⟨wear-(something)-on⟩ your head with that hat'<sup>16</sup>. Such sentences illustrate how Mwotlap SVCS can involve a complete reorganization of argument structure and semantic roles in the clause<sup>17</sup>.

#### [p] Combination of several verbs, two of which are transitive

Finally, the reader will not be surprised to learn that our corpus shows examples of serial chains with more than one transitive verb.

- (53) *Key* ⟨*lañ mat veteg*⟩   *hōw*   *nō-lōmgep*   *en*,  
      3PL   AO:beat die   (leave)   down   ART-boy   ANA  
      *tō*   *qeleglen̄*.  
      then   AO:disappear:DUP  
      'They got rid of the boy by beating him to death, and escaped.'

This example apparently brings about a new, complex formula, in which the same argument (*y* = 'the boy') is simultaneously the object of  $V_1$ , the subject of  $V_2$ , and again the object of  $V_3$ . But as usual, we find no formal clue whatsoever which would encode the internal structure of the macro-

verb. How can the hearer cope with such complexity, and consistently assign the right semantic role to the right argument?

As was stated in [m], this kind of multi-verbal combinations can in fact always be analysed as the result of several binary expansions, starting from the head. It is then easy to retrieve the patterns we established earlier in this study:

- rule [j]   *x-beat-y* + *y-die*   → *x-[kill]-y*
- rule [n]   *x-[kill]-y* + *x-leave-y*   → *x-[get rid of]-y*

#### 3.4. Summary: Few rules, strict rules

The function of Mwotlap macro-verbs is to define a meaningful representation of a single action, performed by a given subject at a particular point in time, exactly the same way as a simple verb would do. This is why, far from linking together any two predicates the same way as would do coordination, Mwotlap serial strategies obey strict rules regarding the semantic compatibility between its components, as well as their syntactic organization. We have shown it was possible to reduce the high diversity of all SVC instances to a finite number of thirteen regular patterns, each one being easily formalized. These patterns all consist in observing how the argument structure of a verb  $V_1$  and a verb  $V_2$  regularly combine, in order –for the linguist as well as the speaker– to calculate what the argument structure of the resulting 'macro-verb' will be.

All these rules<sup>18</sup> are summarized in the following chart, each one corresponding to a single square. In bold, italic letters we represent those cases which are most frequent or productive in Mwotlap discourse, the other ones being much rarer. The two rows correspond to the valency (intransitive vs transitive) of  $V_1$ ; each column informs on the status of  $V_2$ . Notice that the use of letter *z* is only relevant when it contrasts with *y* (= two distinct applicants for one object slot): hence the shaded squares.

Combination of intransitive and/or transitive verbs:  
 Effects upon the distribution of arguments

	<i>x-V<sub>2</sub></i>	<i>ø-V<sub>2</sub></i>	<i>=V<sub>2</sub></i>	<i>y-V<sub>2</sub></i>	<i>x-V<sub>2</sub>-y</i>	<i>z-V<sub>2</sub></i>	<i>x-V<sub>2</sub>-z</i>
<i>x-V<sub>1</sub></i>	<b><i>x-[VP]</i></b>	<i>x-[VP]</i>	<b><i>x-[VP]</i></b>	<i>x-[VP]-y</i>	<b><i>x-[VP]-y</i></b>		
<i>x-V<sub>1</sub>-y</i>	<i>x-[VP]-y</i>	<i>x-[VP]-y</i>	<b><i>x-[VP]-y</i></b>	<i>x-[VP]-y</i>	<i>x-[VP]-y</i>	<i>x-[VP]-z</i>	<i>x-[VP]-z</i>

While detailed discussions have already been given throughout the preceding pages, we will only add a few general comments to this chart. The combination of two verbs in Mwotlap follow a small number of very regular principles, some of which are obvious, while others are less so:

- The number of arguments of the macro-verb is never higher than two.
- The number of arguments of the macro-verb is never lower than the sum of distinct arguments associated with the input verbs.
- The subject of the macro-verb is systematically the same as for  $V_1$  (i.e.  $x$ ).
- The object of the macro-verb is any argument other than  $x$ , associated to the input verbs (i.e.  $y$  or  $z$ ).
- In case of conflict for the object position, any argument introduced by  $V_2$  (i.e.  $z$ , either its semantic subject or object) will have priority over the object of  $V_1$ .
- In other words, the macro-verb adopts the primary orientation<sup>19</sup> of  $V_1$ ; its secondary orientation depends mainly on  $V_2$ .

These rules are most probably operated by the speaker, while creating new combinations and building his discourse – but also by the hearer, while endeavouring to associate the right referents with the right syntactic slots.

The high number of attested patterns does not mean that any combination of two verbs will be possible in Mwotlap. First, as we said already, the inventory of the adjunct  $V_2$  is limited, even for the patterns which are regular (e.g. no example of SVC with  $V_2$ = ‘take’, ‘see’, ‘say’, etc.). Secondly, one can observe that several “potential” combinations are impossible in Mwotlap: \*{ $x\text{-}V_1 + y\text{-}V_2\text{-}x$ }; \*{ $x\text{-}V_1 + y\text{-}V_2\text{-}z$ }; \*{ $x\text{-}V_1\text{-}y + z\text{-}V_2\text{-}y$ }; \*{ $x\text{-}V_1\text{-}y + y\text{-}V_2\text{-}z$ }. The last one, in particular, is worthy of notice, since it would have corresponded to a plausible scenario of X acting upon Y, in order for this Y to act upon Z; this corresponds to an “agentive causative” structure (Fr. *factitif*, opp. *causatif*), e.g. { $I_x \text{ command } you_y + you_y \text{ shut the door}_z \rightarrow *I \text{ command-shut...}$ }. This pattern, which is perfectly possible with Core-layer SVCS<sup>20</sup>, is impossible with the serial structures of Mwotlap: subordination structures will be used instead.

In brief, serialization strategies are not so free and random as one could believe at first sight: not only is the position of adjunct restricted to a finite inventory of verbs, but also the combinations themselves are governed by strict principles. A noteworthy paradox is that, whilst all these constraints and limitations tend to restrain haphazard innovations, they also provide efficient tools to invent novel macro-verbs, with a lesser chance of ambi-

guity, and a higher guarantee that the communication process will be successful.

#### 4. Co-lexicalization and categorial change

##### 4.1. Synchronic heterogeneity and language dynamics

The last section of this paper will be concerned with diachronic change and the language dynamics implied by the serial patterns of Mwotlap. Indeed, one would get too simplistic an idea of these structures if they were to be described as a homogeneous, entirely productive set of transformative rules, applying to any lexical unit in synchrony; and conversely, it would also be erroneous to point to the other extreme, as if every ‘macro-verb’ were already lexicalized and learnt as such by the speakers.

After observing the way these SVCS work in Mwotlap, it seems much more accurate to synthesize these two possible interpretations. Serial verbs in Mwotlap are better defined as a heterogeneous linguistic device, involving several layers in the lexicon, and different depths in time:

- Many combinations are “already lexicalized” from the speaker’s point of view, in such a way that the behaviour of the bundle cannot be unambiguously derived from the meaning of its components. These combinations are learnt as they are, as would be the case for any lexeme or idiom – cf. Eng. *give up*.
  - e.g. *mat mōl* ‘faint’ < *mat* ‘die’ + *mōl* ‘return’.
- A great number of combinations, despite being already there in the language, remain relatively transparent to the hearer, i.e. could be interpreted as the “free” association of a Verb (known separately) + an adjunct (whose meaning is easily reconstructible, and thus ready to be reused) – cf. Eng. *climb up*.
  - e.g. *yow tig* ‘jump or dive while remaining upright’ < *yow* ‘jump’ + *tig* ‘stand’.
- Out of attested combinations, especially the transparent ones, the speaker draws a set of rules. These emerging principles can be syntactic, as we saw earlier, or semantic; they may be general laws governing all serial structures, or more specific rules associated with a particular subset of verbs (e.g. verbs of movement, of speech...), or even with a single lexeme.

- e.g. The lexeme *mat* ‘die, dead’ can be used as an adjunct:
- . combined with a transitive, agentive verb, it will mean that the object dies; thus the macro-verb refers to a way of killing:  
e.g. *ñit mat* /bite+die/ ‘x bite to death y’.
- . combined with an intransitive, active verb, it will mean that the subject metaphorically kills himself in doing an action, i.e. performs it in an intense manner (jocular use); the subject will be repeated as an object of the same verb ([b]):  
e.g. *lak mat* /dance+die/ ‘x dance furiously x’.
- . combined with an intransitive, non-active verb, *mat* is reduplicated, and adds the meaning ‘quietly’; the result is intransitive:  
e.g. *hag matmat* /sit+die:DUP/ ‘x sit still’.
- A certain degree of freedom is allowed within the limits of (and also thanks to) these rules. Far from damping down linguistic creativity, the high specificity of these constraints, both in syntactic and semantic terms, makes it easy for the speaker to create – and for the hearer to understand – novel Verb-Adjunct combinations.
  - e.g. Regarding *mat* ‘die, dead’, any new combination  $\langle V_1 + mat \rangle$  can easily be associated with one of its already attested meanings:
 

<i>woh mat</i> /hit+die/	could be created for ‘x hit to death y’
<i>teñ mat</i> /cry+die/	could be created for ‘x cry hard x’
<i>tag matmat</i> /stoop+die:DUP/	could be created for ‘x stoop still’.

These novel combinations will in turn take part in the definition of slightly different rules, which will be the cause of syntactic changes and semantic shifts through time. The remainder of this paper will illustrate the two evolution paths most commonly found in Mwotlap:

- the co-lexicalization of serial verbs, close to lexical compounding;
- the emergence of productive adjuncts.

#### 4.2. The lexicalization of macro-verbs

The tendency for serialized verbs to co-lexicalize, i.e. eventually behave like a single lexeme, has long been acknowledged (e.g. Crowley 1987: 61). This is a natural phenomenon, especially when one realizes that languages are more than just a list of isolate lexemes, and always involve a high number of idioms, phrases or even whole clauses as part of their resources (Pawley & Syder 1983). Indeed, it would be clearly artificial to see all

macro-verbs as the analysable combination, performed *in situ* by the speaker, of two autonomous lexemes: in the same way as *show off* or *give up* have become opaque to the speakers of English and are just memorized as a single unit, SVCs in Mwotlap also consist massively in “fixed” combinations that owe nothing to the creativity of the modern speaker.

Although this is an informal figure, we would personally estimate that in one day of linguistic interaction in Mwotlap, about 90 per cent of  $\langle \text{Verb-Adjunct} \rangle$  combinations are nothing but the repetition of combinations already heard by the speaker – whereas only 10 per cent(?) are novel associations which are created by applying productive rules on this matter<sup>21</sup>. Among all these ‘already heard’ combinations, many will eventually carry some semes which are not necessarily present or predictable from its components. A small selection of them is given below:

- (54) *ak* ‘make’ + *goy* ‘(cover, obstruct... )’ [see fn.17]
  - *ak goy* ‘reserve s.th. for o.s., put a taboo on s.th.; reserve (a woman) as o.’s future wife, get engaged to’
- (55) *dēm* ‘think’ + *liwo* ‘big’
  - *dēm liwo* ‘consider with respect; be worried about; boast, show off’
- (56) *dēm* ‘think’ + *veteg* ‘leave’
  - *dēm veteg* ‘forget on purpose; give up, forgive; omit’
- (57) *tog* ‘live somewhere, stay for several days’ + *qōñ* ‘be night’ [see 21]
  - *tog qōñ* ‘go for a picnic, spend the afternoon somewhere’
- (58) *tot* ‘chop’ + *gal* ‘lie, tell lies’
  - *tot gal* ‘notch wood; sculpt, carve an image; portray, take a photo or video of’.

Obviously, such semantic shifts make it necessary to consider these so-called “serial verbs” as single lexemes (written in one word?), and thus treat them as separate entries in a dictionary of Mwotlap (in preparation).

The question arises whether serial verbs in Mwotlap, or more generally  $\langle \text{Verb-Adjunct} \rangle$  combinations, should be considered as a kind of verb compounding. We do not exclude this interpretation, considering it at least to be better – as far as Mwotlap is concerned – than the purely serializing interpretation [see fn.3]. Nevertheless, three reasons refrain us from speaking of lexical compounding:

- 1) *A phonological criterion*: Whatever their semantic evolution, Verb and Adjunct are always treated as two distinct phonological words, as is

shown by the syllabic template mapping on each element separately<sup>22</sup>. Compare *ni-* ‘prefix 3SG’ + *v(e)teg* ‘leave’ → *ni-pteg* (‘he leaves’); but *hō* ‘paddle’ + *v(e)teg* ‘leave’ → *hō v̥eteg* (\**hō-pteg*) ‘paddle away from, quit’.<sup>23</sup>

2) *A syntactic criterion*: Under certain conditions, it happens that some other adjunct may insert between the two elements of a “compound”.

3) *A semantic criterion*: As the reader will have noticed throughout this paper, it is often difficult to draw the line between those cases in which it is ‘clear’ that *(Verb + Adjunct)* form a single lexical unit, and those cases in which it is ‘clear’ that it consists of two autonomous notions.

In order to describe the facts of Mwotlap, it is not necessary to follow the old-fashioned definition of the ‘lexeme’ as coinciding with a single word; what we have here are whole phrases *(Verb + Adjunct)* which are more or less (co-)lexicalized as a fixed formula.

#### 4.3. Intransitive vs transitive adjuncts

The second diachronic path which is often followed by Mwotlap SVCS, is for the second verb V<sub>2</sub> either to undergo a radical semantic shift as opposed to its use as a verbal head, or sometimes completely lose the latter, and specialize as a pure adjunct.

##### 4.3.1. Semantic split Verb / Adjunct

A good example of a semantic shift affecting a serialized verb, is provided by the verb *tēy*: as a main verb, it normally means ‘hold, have in o.’s hands’. Used as an adjunct with a movement verb V<sub>1</sub>, *tēy* will help express such meanings as ‘bring, take away (s.th. somewhere)’. The basic meaning ‘hold’ is maintained in these examples [see (44)], but widens to cover any value of X moves with Y, including ‘accompany (s.o. somewhere)’:

- (59) *Hiqiyig* <*ni-hō tēy*> *tita!*  
someone AO-paddle hold Mum  
'Someone takes Mum in their canoe!'

This combination is not restricted to movement verbs; with other verbs too, *tēy* takes a general comitative meaning:

- (60) *No* <*ta-lak tēy qiyig*> *na-savat.*  
1SG FUT-dance hold IMM ART-shoes  
'I will dance with my shoes on.' (= on my feet, not \*in my hands)

Finally, the same verb *tēy* sometimes encodes a more abstract relation *f(x,y)*, with no trace of the original, lexical meaning ‘hold in o.’s hands’:

- (61) *Kēy* <*ne-mlē*> → *Kēy* <*ne-mlē tēy*> *na-mwumwu namuy.*  
3PL STA-slow 3PL STA-slow hold ART-work  
their

'They are slow.' 'They are slow AT their work.'

The kind of ‘semantic bleaching’ involved here suggests the possibility of a shift from a lexical, semantically specific use of *tēy* (*x holds y in x's hands*), to a ‘grammatical use’ as a “transitivizer morpheme” (by performing an action V<sub>1</sub>, *x affects y in some manner*)<sup>24</sup>.

##### 4.3.2. Emergence of pure adjuncts

Ultimately, the two uses (as a verb ≠ as an adjunct) of the same lexeme will be felt to belong to two distinct linguistic units; and the decay which may affect, say, the verb will not affect the homophonous adjunct. As a consequence, the lexeme eventually leaves the category of verbs, and resembles the ‘pure adjuncts’ already existing in the language [cf. *tiwag* in ex. (4)]. The same evolutionary path was described for Paameese nuclear-layer SVCS, by Crowley (1982: 167):

The meanings of these forms as verb phrase heads and as adjuncts diverged so widely that they were no longer felt to represent the same morpheme. Subsequently, the normal processes of lexical loss and replacement resulted in the situation where some of these forms as verb phrase heads ceased to exist.

From the syntactic point of view, new adjuncts proceed either from an intransitive or from a transitive (ex-)verb. In the first case, the result will be the same as those dozens of adjuncts which have a purely semantic function, and no effect on valency. Although they do not behave as verbs in synchrony, it would be perfectly plausible that former verbs are the source of modern adjuncts like *qēt* ‘[do s.th.] completely’, *vatag* ‘already’, *ēgē* ‘hastily’, *tēqēl* ‘[go] down’, *woy* ‘[split] lengthwise’...<sup>25</sup> In other cases, however, an intransitive adjunct may originate in a word class other than

verbs, e.g. *ēwē* ‘just’ < adjective ‘good’ [ex. (47)]; or *lō* ‘out’ < POc \**lua*, apparently an adjunct since early stages of Oceanic.

In the second case, a transitive verb has become what may be named a “transitive adjunct”. Contrary to the latter, these adjuncts have the power to alter the argument structure of a verb *V*<sub>1</sub>, either by increasing its valency if *V*<sub>1</sub> is intransitive, or by modifying its secondary orientation if it is already transitive (following the same rules as in the table p. 133). For example, the adjuncts *sas* ‘(find)’ and *day* ‘(await)’, although they are not verbs in synchrony, not only change the meaning of the preceding verb, but also transitivize it:

- (62) *Kōyō* <*ma-van SAS*> *ni-tilto*.  
 2DU PFT-go (find) ART-egg  
 ‘As they were walking, they found an egg.’
- (63) *Nok* <*tigtig DAY*> *nēk*.  
 1SG AO:stand:DUP (await) 2SG  
 ‘I was (standing) waiting for you!’

Sometimes, a lexeme is already totally specialized as a transitive adjunct in younger people’s speech – but suddenly some archaic or literary phrase uses it as a VP-head, betraying its true origin. This is the case with *yak*, which is employed as an adjunct ‘[take, go...] away’ in about ninety-nine per cent of its uses, but appears as a verb ‘pick up’ in a ritual phrase connected with marriage (*yak nē-sēm* ‘[the bride’s father] picks up the money [given by the groom]’).

Similarly, the word *v(e)teg* which we have been glossing ‘leave’ is very seldom used as a verb; most of the time it forms a transitive adjunct, expressing a physical or abstract separation between the subject and the object [ex. (45)-(50)-(53)-(57)]. This includes the coding of comparison, a metaphorical extension of the idea of separation:

- (64) *Kē* <*nē-mnay VETEG*> *nēk*.  
 3SG STA-clever (leave) 2SG  
 ‘He is cleverer than you.’

As is shown in this last example, the so-called “transitivizing” function of some adjuncts does not necessarily affect verbs as such, but also adjectives or other categories; it would be better defined in terms of logical relations (Lemaréchal 1998), as a device allowing combining a first unary predicate –here the adjective *f(x)*= ‘clever’– to a binary, relational predicate *f(x,y)*= ‘be superior to’. This is how the comparative structure, *par*

*excellence* relational, happens to be coded by a “transitive adjunct”, originally a verb.

These transitive ex-verbs look as if they had transformed, one may argue, into prepositions: but what is true for other languages does not fit the structures of Mwotlap, where the strong VP-final boundary preserved *V*<sub>2</sub> from becoming a preposition<sup>26</sup>. One could then think of another possible category, i.e. grammaticalizing into an applicative morpheme<sup>27</sup>; however, contrary to applicatives, we have already seen that these adjuncts are capable of modifying the diathetic orientation of the verb *V*<sub>1</sub>, in such a way that the direct complement of the macro-verb eventually takes over the semantic role of patient.

Consequently, the new ‘grammatical unit’ arising from the diachronic specialization of transitive verbs in the adjunct-position, should no longer be described as a verb any more (*contra* Crowley 1987: 61), nor as a preposition or an applicative. Rather, the most reasonable analysis would certainly consist in sticking to the “vernacular” category of (intransitive vs transitive) adjunct, to which are associated certain syntactic and semantic properties, different from other word classes. Although this approach seems to partially hamper cross-linguistic comparison, it has the paramount advantage of allowing a deeper understanding of the synchronic structures, as well as the historical evolution, of Mwotlap grammar. Thanks to this methodological caution, it becomes obvious why some nouns, some adjectives, some verbs (etc.) seem to be merging into a single category, obeying the same rules, whenever they follow the head of a VP. This is the only way a vernacular-oriented analysis of the language becomes feasible, as a necessary requirement before any cross-linguistic statement is made.

## 5. Conclusion: Serial verbs or Adjuncts?

Although, at first sight, Mwotlap is a good example of what we expect a “serializing language” to be, the present paper showed several reasons why such a tag would mislead linguistic analysis rather than help it. The case where a verbal head *V*<sub>1</sub> is indeed followed by one or more verb roots, rather than being interpreted as a serial verb construction, appears to be merely a subcase of a more general and productive pattern in this language, i.e. the syntactic function of adjunct. As far as these adjuncts are concerned, the only grammatical boundary that really matters is not so much the

distinction between parts of speech (adjectives, verbs...), but the contrast between unary and binary predicates:

Unary vs binary predicates in the adjunct position

Type of predicate	Argument slots	Syntactic categories
Unary predicate	1 slot: $f(x)$	<i>adjectives + nouns + intransitive verbs + intransitive pure-adjuncts</i>
Binary predicate	2 slots: $f(x,y)$	<i>transitive verbs + transitive pure-adjuncts</i>

While unary predicates have little effect upon the argument structure of the macro-verb, binary adjuncts do influence its syntactic behaviour, following complex rules which were detailed here. The fact that some verbs, along with other word classes, are taking part in this mechanism, is indeed worth of notice; but the quest for data on “serial verbs” as such should not obscure the main pivot of the system, which is not so much the verb as a lexical category, but rather the adjunct as a syntactic function in the clause.

### Orthographic conventions and abbreviations.

The spelling conventions adopted for Mwotlap include the following:

$\bar{e} = [i]$ ;  $\bar{o} = [u]$ ;  $g = [y]$ ;  $b = [^m b]$ ;  $d = [^n d]$ ;  $q = [^k p^w]$ ;  $\bar{m} = [^n m^w]$ ;  $\bar{n} = [n]$ .

The abbreviations used in literal translations include:

ANA	anaphoric marker	PRD	predicative
AO	Aorist	PRET	Preterite
ART	article	PRIOR	Prioritive
DU	dual	PROH	Prohibitive
DUP	reduplicated form	PRSP	Prospective
IMM	Immediate future	STA	Stative
NEG	negation	1EX	first exclusive
PFT	Perfect	1IN	first inclusive
POT	Potential		

### Notes

1. In this regard, Mwotlap contradicts the tendency proposed by Foley and Olson (1985) that an SVO language should make use of Core juncture rather than Nuclear juncture. Other SVO languages, like Paameese (Crowley 1987:82) and Lewo (Early 1993:88), have also been mentioned in this respect.
2. This description of  $V_2$  as being essentially a “modifier” to  $V_1$  should be taken for what it is: a first attempt to describe Mwotlap serial verbs in broad, non-technical terms. The following pages will show that the function of  $V_2$  is obviously more complex, and cannot be reduced to a role of lexical specification (see for example the case of causative serialization). The relevant point at this stage of our presentation is to underline the syntactic asymmetry between  $V_1$  (the head) and  $V_2$  (the adjunct).
3. By preferring the ‘adjunct’ interpretation rather than the ‘serializing’ one, we support – at least for Mwotlap – the analysis of Early (1993: 80-81) and even Crowley (1982:166), *contra* Crowley (1987:59).
4. Foley (1986: 117), from Olson (1981). See also the Kalam examples cited by Senft (this volume), after Pawley (1993: 95).
5. A sentence like this one clearly refers to several distinct actions (‘kill, cut up, eat’...). Saying this does not contradict the hypothesis, often formulated (Givón 1991a; Durie 1997: 291), that globally only one event is involved.
6. The clear presence of these pauses, as well as coordinators, makes it impossible to talk about Core-layer serialization. The only case which shows Core-layer SVC is reserved to purpose clauses in an Irrealis context, e.g. following a general statement, a prospective clause or an order (Aorist): *Lep me nē-bē nok in*. ‘Give me some water [so that] I drink’ (François 2003: 187). Even though the absence of pause between both clauses strongly recalls Core-layer serialization, it must be noted that this structure is functionally very limited.
7. Likewise, Ross (1998:35) chooses to talk about “adjectival verbs” rather than “adjectives” for Proto Oceanic. François (2001) discusses in detail the issue of Mwotlap parts of speech, including the adjective *vs* verb contrast.
8. A similar phenomenon occurs in English: compare *You hit him to death* (with transitive *hit*), and *You danced yourself to death* (with intransitive *dance*).
9. This corresponds to what Crowley (1987: 40, 49) labelled ‘ambient serialization’ (e.g. /I hit you it was hard/ = ‘I hit you hard’); but his description of  $V_2$  as a “general” predication, in our opinion, somewhat lacks precision.
10. The corresponding statements make use of an “impersonal” subject *mahē* (‘place’) for this kind of weather sentence: *Mahē mō-qōñ* ‘It is night’ [*lit.* The place is night]. That *mahē* is an empty argument, not a true one, is proven

- precisely by the impossibility of finding it as an object of a  $\langle V_1-V_2 \rangle$  bundle: thus  $\$[d] \neq \$[e]$ .
11. Further evidence against this general prediction will appear in §[k].
  12. The verb S.P.R. is a loanword from pidgin bislama S.P.R. [Espiari], which stands for *Sperem Pablik Rot*, lit. ‘hit the public road’, i.e. ‘roam, wander aimlessly all day long’. This has become a full verb in Mwotlap, including the possibility of root reduplication, hence S-S.P.R. [EsEspiari].
  13. Mwotlap has lost the POc causative prefix \*pa/ka-, except for two residual, unanalysable, verbal lexemes *wot* ‘be born’ > *vawot* ‘give birth to’; *ēh* ‘live’ > *vaēh* ‘save s.o.’s life, rescue, cure’.
  14. For other “potential” formulas, the risk of semantic ambiguity of a serialized VP would be so high that subordination is always the rule: see the case of {x- $V_1-y + y-V_2-z$ } p.134.
  15. The notion of relational seme appears to be particularly fertile in the theory developed by Alain Lemaréchal, regarding the semantic and syntactic mechanisms of verb serialization (1998: 206-233).
  16. Very similar sentences are reported for neighbouring Mota: *saru GORO natarapeñā mun o siopa* ‘clothe over your body with a garment’ (Codrington 1885: 293).
  17. The last example, along with many others, is commented in detail in a paper (François 2000 b) dedicated to the numerous valency-effects of this adjunct (not verb) *goy* [ $V+goy$ ] = ‘cover, obstruct, occupy, forbid, react, disturb, dominate, reserve, fetch, protect, assist...’ (< POc \**koro* ‘to surround; fortified village’)] and to its interest for syntactic and semantic theory.
  18. All, except [b] – which is exceptional anyway – plus the cases of combinations involving more than two verbs ([f], [m], [p]); we showed that these could always be analysed into smaller, binary combinations.
  19. The notion of diathetic orientation of the verb, “primary orientation” towards the subject and “secondary orientation” towards the object, is developed by Lemaréchal (1989).
  20. For example, see François (2002: 112 99) for Araki: *Nam vadai-a ni-a co les moli* ‘(lit.) I told him (that) he should see the chief’.
  21. This kind of approximation would be very difficult to check. Among other things, this issue will also depend on the speaker’s personality; the more easily one breaks new ground in forging SVCS, the better reputation he will get as a language *connoisseur*.
  22. This point is more detailed in François (to appear); for similar remarks about Paamese, see Crowley (1987: 60).
  23. In the terminology proposed by Durie (1997: 302-303), Mwotlap SVCS would thus be defined as [+ contiguous] because  $V_2$  immediately follows  $V_1$ , but

- [- incorporating], since the two verb roots remain as two distinct phonological words.
24. Although the verb ‘take’ (Mtp *lep*) must be distinguished from ‘hold’ (*tēy*), the situation here described strongly recalls the evolution of Mandarin *ba* into an Object marker (Li and Thompson 1974c); closer to us, see the New Caledonian languages described by F. Ozanne-Rivierre (this volume).
  25. This general remark sometimes corresponds to a mere hypothesis, and sometimes is confirmed by further etymological or dialectological research. Although such forms as *qēt*, *tēqēl*, *day* are only used as pure adjuncts in Mwotlap, their cognate forms in Mwesen (Vanua-lava I., 8 speakers; pers. data) are still verbs: *qēt* ‘finish’, *tēqēl* ‘go down’, *nar* ‘await’. Similarly, *woy* corresponds to a verb in most Vanuatu languages <(Proto NCV) \**vora* ‘break, divide, split’ (Clark 2000), etc.
  26. There are a few exceptions to this “impossible evolution”, since a few adjuncts have historically crossed over the VP (right) boundary. For instance, the Proto Oceanic verb \**suri* [glossed ‘prepositional verb: allative’ by Ross (1988)] appears in Mwotlap not as a verb, but as a transitive adjunct *hiy* ‘(do  $V_1$ ) having s.th. in mind’, thus still VP-internal; on the other hand, it has also grammaticalized as a preposition, outside the VP, coding for Dative.
  27. An Applicative morpheme is a VP-internal marker modifying the diathesis of a verb, in such a way that what could have been coded as an oblique complement is construed directly, as if it were an object, but it does not eliminate the other object. Applicatives are a kind of ‘incorporated preposition’: cf. Guinea Fula *En habb-ay ledde* ‘we shall bind the wood’ → *En habb-ir-ay ledde boggol* ‘(lit.) we shall bind-with the wood a cord’ (Labatut n.d.: 126). On Applicatives, see Palmer (1994), Lemaréchal (1998: 189).



## DIVERSITÉ DES PRÉDICATS NON VERBAUX DANS QUELQUES LANGUES OCÉANIENNES

### Abstract

*It is a common feature among Oceanic languages, that most parts of speech (especially nouns) are capable of becoming the head of a predicate, without needing any copula. Our aim here is to present the principles and properties of this so-called ‘generalised predicativeness’; in doing so, we shall pay special attention to its direct and indirect effects upon the whole system of a language: the need for a language-specific definition of syntactic categories; the semantics of predication and tense-aspect-mood; the internal economy of the lexicon, as well as the issue of derivation. After presenting the structures of Mwotlap (Vanuatu), this paper will end with an overview of other Oceanic languages.*

Les langues océaniennes, au nombre d'environ cinq cents, se répartissent géographiquement entre la Micronésie, la Mélanésie et la Polynésie; elles forment ensemble la branche la plus orientale de la grande famille austronésienne<sup>1</sup>. Même s'il faut se garder des généralisations s'agissant d'une famille aussi nombreuse et encore partiellement méconnue, on peut affirmer que la grande majorité des langues d'Océanie se caractérisent par l'absence de copule, et par une forme assez extrême d'omniprédictativité. Dans certains cas, les noms prédicatifs entrent dans des structures syntaxiques qui leur sont propres et les distinguent des verbes; dans d'autres cas, il arrive au contraire que les noms se comportent largement comme les verbes au regard de la prédication, au point qu'il devienne parfois nécessaire d'affiner l'argumentation pour distinguer les uns des autres.

Nous examinerons d'abord la diversité des prédicats non-verbaux dans une langue de Mélanésie, le mwotlap. Dans un deuxième temps, nous comparerons brièvement ces structures avec d'autres configurations syntaxiques au sein de la même famille océanienne, à travers les faits de quelques langues de Mélanésie et de Polynésie. Dans tous les cas, nous chercherons à observer les implications de ces prédicats non-verbaux, sur des plans aussi variés que la syntaxe de l'énoncé, l'organisation du lexique ou la valeur de la prédication en termes sémantiques et cognitifs.

1. Cet article trouve son origine dans un exposé à deux voix, présenté par Claire Moyse-Faurie et moi-même, lors de la Journée d'études 2003 de la SLP. S'agissant de la version écrite, nous avons choisi d'en consacrer l'essentiel à une langue (le mwotlap), en renvoyant par ailleurs le lecteur aux parutions que C. Moyse-Faurie a déjà publiées sur ce sujet, concernant notamment les langues polynésiennes (voir §2).

## 1. Les structures prédictives du mwotlap

Le mwotlap est parlé par environ 1800 locuteurs au nord du Vanuatu, principalement dans l'île de Motalava (François 2001 *b* ; 2003). Dans cette langue, la répartition des lexèmes en classes syntaxiques distinctes ne pose pas de problèmes majeurs, du moins en dehors des contextes prédictifs: ainsi, on identifie un radical comme un nom par sa capacité à former la tête d'un syntagme actanciel (précédé de l'article nominal *na-* s'il réfère à un non-humain); les adjectifs sont définis par leur capacité à figurer en position d'épithète directement après le nom; les verbes, enfin, ne sont compatibles avec aucune de ces deux positions, et ne figurent qu'à l'intérieur des syntagmes prédictifs.

Pourtant, la netteté de cette distinction s'estompe quelque peu dès lors que l'on considère la syntaxe du prédicat. En effet, les noms et les adjectifs partagent avec les verbes la capacité de former directement un prédicat, sans copule de type *être*. Dans certaines structures, l'équivalence fonctionnelle de ces trois classes de mots est si nette, qu'il serait tentant de les faire entrer toutes les trois dans une unique «macro-catégorie», celle des lexèmes prédictables<sup>2</sup>.

Nous proposons d'examiner plus en détail, dans un premier temps, les divers types de prédicats en mwotlap, et leurs liens avec les catégories syntaxiques de la langue (§1.1). Dans un deuxième temps, nous observerons un cas particulièrement net de convergence fonctionnelle entre noms, adjectifs et verbes: la combinaison avec les marques aspecto-temporelles (§1.2). Enfin, nous discuterons des restrictions et limites qui permettent de définir les prédicats non-verbaux (§1.3).

### 1.1. Les types de prédicat

La syntaxe du mwotlap met en jeu essentiellement quatre structures prédictives distinctes. Dans tous les cas, l'ordre de l'énoncé est Sujet-Prédicat, S.V.O. pour les verbes transitifs. La fonction des arguments (pronoms personnels, syntagmes nominaux) est exclusivement marquée par leur position dans la chaîne, sans marque de cas ni d'accord sur le prédicat.

#### 1.1.1 Prédicat marqué en Temps-Aspect-Mode (TAM)

La structure prédictive la plus fréquente en mwotlap est celle qui associe une tête lexicale à une marque de temps-aspect-mode<sup>3</sup>: par commodité, nous appellerons cette structure *prédicat TAM*. Pour tous les verbes, la seule manière d'accéder à la prédictivité est de constituer la tête d'un prédicat TAM<sup>4</sup>:

2. Voir chez Lemaréchal (1989: 27 sq.) la notion de «super-partie du discours» — en l'occurrence, celle qu'il nomme Qualificatifs (nom, adjectif, verbe).

3. Les marques TAM forment en mwotlap un paradigme unique de vingt-six membres; le système est détaillé dans François (2003).

4. Les conventions orthographiques pour le mwotlap incluent: *ē* [i]; *ō* [u]; *g* [y]; *b* [mb]; *d* [nd]; *q* [kpʷ]; *m̄* [jmʷ]; *n̄* [nj]. Par ailleurs, chaque fois que nous le jugerons utile, nous indiquerons par des crochets pointus (...) les limites du syntagme prédictif.

- (1) *Bulsal mino* ⟨may dēñ⟩ *me*.  
ami mon ACP arriver ici  
'Mon ami est déjà arrivé.'

\**Bulsal mino dēñ me.*

Du fait de l'absence de copule, les adjectifs se combinent directement aux marques TAM, dans les mêmes conditions. En d'autres termes, au regard de la prédication, les adjectifs (excepté ceux du §1.1.4) se comportent en tous points comme des verbes statifs. La marque TAM qui accompagne généralement les adjectifs est celle du Statif<sup>5</sup> *nV*:

- (2) *Bulsal mino* ⟨nē-mnay⟩.  
ami mon STA-intelligent  
'Mon ami est intelligent.'

\**Bulsal mino mēnay.*

D'une certaine façon, c'est la marque TAM qui permet aux deux catégories VERBE et ADJECTIF de remplir la fonction prédictive. Nous verrons bientôt (§1.2) que les noms sont également compatibles avec cette structure aspecto-modale, sous certaines conditions sémantiques; cependant, s'agissant des noms, la tournure la plus fréquente consiste en un prédicat direct, sans marque TAM.

#### 1.1.2 Prédicat nominal direct

Un autre type de prédicat consiste à identifier un sujet X en le superposant à une entité Y déjà construite dans le discours: il s'agit d'un prédicat équatif, que l'on peut glosser *X c'est Y*. Ces prédicats se composent d'un simple syntagme nominal, en tous points similaire à un syntagme actanciel. En conséquence, les énoncés que l'on obtient prennent souvent la forme d'une juxtaposition de deux syntagmes actanciels, de type { *X<sub>SN</sub> Y<sub>SN</sub>* }:

- (3) *N-ēñ gōh* ⟨n-ēñ mino⟩.  
ART-maison cette ART-maison ma  
'Cette maison, c'est ma maison.'

Par ailleurs, le mwotlap n'opère pas de différence formelle entre les prédicats équatifs stricto sensu (*X c'est [le] Y*) et les prédicats d'inclusion, lesquels consistent à simplement caractériser le sujet X par une propriété nominale définitoire (*X est [un] Y*). La structure utilisée ici est également celle d'un prédicat nominal direct, sans copule ni marque TAM<sup>6</sup>:

- (4) *Imam mino* ⟨tēytēybē⟩.  
père mon guérisseur  
'Mon père est médecin.'

5. Le statif est employé chaque fois qu'il s'agit de prédiquer du sujet une qualité stable, à un moment donné (passé, présent, futur). En ce qui concerne la combinaison des adjectifs avec d'autres marques TAM, voir §1.2.3.

6. Quant à l'article *na-*, il est normalement réservé aux noms sémantiquement non-humains. Sa présence — comme en (3) — ou son absence — comme en (4) — dépendent donc du sémantisme du nom en jeu, pas du type de prédicat.

Ces deux types de prédictats, que nous réunirons sous le terme générique de prédictat nominal direct, ne peuvent avoir comme tête que des noms, jamais des adjectifs ou des verbes.

Lorsque le sujet est repris par anaphore zéro, il n'est pas rare que l'énoncé coïncide avec un prédictat nominal, que seules les marques intégratives (frontières prosodiques de proposition, intonation assertive) distinguent d'un SN actancial:

- (5) *⟨Na-naw⟩.*  
ART-eau.salée  
'C'est de l'eau de mer.'

### 1.1.3 Prédicat locatif

Les prédictats nominaux directs [X ⟨Y⟩ = X est Y] que nous venons de voir ne doivent pas être confondus avec une structure ressemblante, dans laquelle le prédictat apparaît également construit directement (sans copule), mais qui consiste à localiser le sujet dans l'espace [X ⟨Y⟩ = X se trouve en Y]. Sémantiquement parlant, ces prédictats locatifs se reconnaissent à l'absence de coréférence entre X et Y: Y est un localisateur pour X.

En (6), le toponyme *Iñglan* est un prédictat équatif (type X est Y); en (6'), le même mot forme un prédictat locatif, et ce, sans la médiation d'aucun autre morphème:

- (6) *Na-pnō mino ⟨Iñglan⟩.*  
ART-pays mon Angleterre  
'Mon pays, c'est l'Angleterre.'
- (6') *Bulsal mino ⟨Iñglan⟩.*  
ami mon Angleterre  
'Mon ami est/se trouve en Angleterre.'

Le contraste entre les deux tournures peut être mis en évidence formellement, si la tête du prédictat est constituée non par un nom propre de lieu, mais par un nom commun, ex. *ēm* 'maison'. Alors que ce nom en prédictat équatif portera l'article (comme en (3) ci-dessus) il ne pourra former un prédictat locatif qu'en étant marqué par la préposition locative *IV-*:

- (7) *Bulsal mino ⟨l-ēm mino⟩.* \**Bulsal mino n-ēm mino.*  
ami mon dans-maison ma  
'Mon ami se trouve chez moi.'

La classe des locatifs constitue d'ailleurs, en mwotlap, une catégorie distincte de celle des noms — en dépit de quelques chevauchements, illustrés en (6)-(6'); elle inclut les toponymes (ex. *Iñglan*) et de nombreux adverbes (ex. *alge* 'en haut')<sup>7</sup>.

7. Ainsi, il serait plus exact de présenter le préfixe *IV-* non comme une préposition, mais comme un «translatif», au sens de Tesnière (1953) et Lemaréchal (1989): *IV-* sert à translater n'importe quel nom commun en locatif, afin de le faire accéder au même éventail de fonc-

Au passage, on peut ranger dans cette même catégorie le prédictat existentiel affirmatif *aē* 'il y a', qui résulte lui-même d'une extension d'emploi de l'adverbe anaphorique inanimé *aē* 'y, là, avec/pour cela...'. Si le sujet est lui-même marqué en possession, la tournure *Mon X existe* traduira notre verbe *avoir*:

- (8) *Bulsal mino ⟨aē⟩.*  
ami mon y  
a) 'Mon ami s'y trouve (~ s'y trouvait).'  
b) (Mon ami existe) = 'J'ai un ami.'

### 1.1.4 Prédicat direct

Enfin, certains éléments sont directement prédictifs, sans être pour autant ni des substantifs, ni des adverbes locatifs. On peut en distinguer quatre sous-types, qui ne concernent chacun qu'un nombre limité de mots.

#### — Les attributs directs

Le mwotlap compte une dizaine de pseudo-adjectifs ou «attributs», qui n'apparaissent essentiellement qu'en position de prédictat, et le font directement (sans marque TAM): *itōk* 'bon'; *namnan* 'parfait'; *yeh* 'loin'; *isqet* 'proche'; *haytēyēh* 'convenable'...

- (9) *Bulsal mino ⟨yeh⟩.*  
ami mon être.lointain  
'Mon ami est loin (d'ici).'

#### — Les numéraux

Les numéraux, lorsqu'ils sont prédictats, le font aussi de façon directe. Ainsi, seules des marques prosodiques permettent de distinguer entre un numéral épithète (ex. *bulsal mino vētēl* 'mes trois amis') et un numéral prédictatif, formant un énoncé complet:

- (10) *Bulsal mino ⟨vētēl⟩.*  
ami mon trois  
(litt. mes amis sont trois) 'J'ai trois amis.'

#### — Les possessifs

Un cas particulier de prédictat direct concerne les prédictats d'appartenance. Pour traduire *X appartient à Y*, on peut bien entendu construire un prédictat équatif avec pour tête un nom suivi d'un possessif: c'est la tournure que nous avons vue en (3). Mais le mwotlap permet également à la marque possessive elle-même de constituer un prédictat direct:

tions syntaxiques (circonstant, thème, prédictat locatif...) que la classe de locatifs directs. Cf. François (2001 b: 164 sqq.).

- (3') *N-ēm gōh <mino>*.  
 ART-maison cette ma  
 'Cette maison est à moi (*litt.* est mienne).'

Cette structure concerne les quatre classificateurs possessifs aliénables de la langue, qu'il s'agisse de *mino* (possession aliénable en général), de *namuk* (possession temporaire), de *nakis* (nourriture) ou de *nemek* (boisson):

- (11) *Na-ṁalṁal gōh <nakis>!*  
 ART-jeune.fille cette Possessif.Comestible:1sg  
 'Cette fille, elle est pour moi!'

#### — Syntagmes et propositions

Pour terminer, il faut noter qu'un prédicat direct peut être constitué par toute une proposition, ou un syntagme prépositionnel [cf. (7)], avec les mêmes valeurs que le français *C'est (que P)...* Le phénomène est particulièrement visible en contexte négatif, car le syntagme prédicatif porte directement la négation (*et... te*). Si la séquence est longue, ladite négation encadre simplement la préposition ou la conjonction, dans une tournure d'ailleurs assez exotique:

- (12) <*Et-qele te na-pnō nōnōm*>. NÉG<sub>1</sub>-comme NÉG<sub>2</sub> ART-pays ton  
 'Ce n'est pas comme ton pays.' [*litt.* Ne comme pas ton pays]
- (13) <*Et-veg te so n-eh itōk*>. NÉG<sub>1</sub>-car NÉG<sub>2</sub> que ART-chanson être.bien  
 'Ce n'est pas parce que la chanson est belle.'  
 [*litt.* Ne car pas la chanson est belle]

#### 1.1.5 Récapitulation

Une langue comme le mwotlap rappelle combien il serait abusif de poser *a priori* un lien privilégié entre la catégorie des verbes et la prédicativité, comme on l'a longtemps fait sous l'influence des langues à copule — notamment européennes; même si ce vieux préjugé a été remis en question grâce, notamment, aux études de Lemaréchal (1989) et Launey (1994) sur l'omniprédicativité, il n'est pas superflu de reprendre la démonstration avec de nouveaux exemples. En mwotlap, le prédicat verbal n'est qu'un type syntaxique parmi d'autres — même si, d'un point de vue statistique, c'est probablement le plus fréquent dans le discours. Pas moins de six classes syntaxiques sont intrinsèquement capables de former un prédicat direct, sans copule ni marque TAM: il s'agit des noms (prédicat équatif ou d'inclusion), des locatifs, des attributs, des numéraux, des classificateurs possessifs, et de certaines propositions. Par ailleurs, les verbes et les adjectifs ne peuvent pas former de prédicat directement, et doivent, pour ce faire, être marqués en temps-aspect-mode.

#### 1.2 Les noms marqués en Temps-Aspect-Mode

Que les noms soient susceptibles de former directement un prédicat direct, comme en (3) ou en (4), est en réalité assez commun d'un point de vue typologique: on trouve des phénomènes similaires en russe, en arabe ou en nahuatl. Mais dans la plupart des langues de ce type, une copule devient nécessaire dès lors que l'énoncé implique des temps, aspects ou modalités distincts du simple présent *realis*: c'est le cas de *byt'* en russe ou de *kazna* en arabe classique (voir d'autres exemples en §2.1.1).

Au contraire, le mwotlap rend ses noms compatibles non seulement avec la fonction prédicative, mais également avec l'ensemble du paradigme des marques TAM. Certes, d'un point de vue statistique, la grande majorité des prédicats nominaux prend la forme directe que nous avons déjà analysée; mais dans le principe, la latitude de combinaison entre noms et temps-aspect-mode est libre et totale. Nous allons examiner plus précisément les conditions d'apparition de ces «noms tamophoriques»<sup>8</sup>.

##### 1.2.1 L'absence de codage temporel

Une première remarque est nécessaire, qui concerne le marquage TAM du mwotlap en général. Les vingt-six morphèmes de ce paradigme mettent en jeu des valeurs sémantiques de nature modale (ex. Futur, Potentiel, Prohibitif...) et/ou aspectuelle (ex. Parfait, Accompli, Aoriste, Statif...); mais la catégorie du temps, au sens de temps déictique référant à l'instant d'énonciation, ne se trouve pas, à proprement parler, grammaticalisée dans cette langue. Ainsi, si l'on reprend quelques-uns des énoncés déjà cités, l'aspect accompli illustré en (1) pourra se calculer par rapport à un repère présent (traduction 'Mon ami *est* déjà arrivé') aussi bien qu'un repère passé ('...*était* déjà arrivé') ou même futur ('...*sera* déjà arrivé'), et ce, sans aucune marque permettant de lever l'ambiguïté; même chose pour le Statif *nV-* en (2) ('il *est/était/sera* intelligent'). Le même principe structural, celui de l'absence de temps grammatical, explique pourquoi tous les prédicats directs que nous avons cités restent ambigus du point de vue de la référence temporelle, qu'il s'agisse d'un pseudo-adjectif de type (9) ['X *est/était/sera* loin'], d'un locatif de type (6')/(8) ['X *s'y trouve/trouvait, j'ai/j'avais X...*'], et ainsi de suite.

On ne s'étonnera donc pas de savoir que le décalage temporel par rapport à la situation d'énonciation (passé, futur) ne constitue pas, en mwotlap, une condition suffisante pour qu'un nom se combine à une marque TAM. Ainsi, alors que l'arabe syro-libanais distingue (Ø) *bayy-e tabizb* 'mon père est médecin' de *kezn bayy-e tabizb* 'mon père était médecin' (Samia Naïm, comm.pers.), ces deux énoncés seront confondus en mwotlap, sous la forme d'un prédicat direct (4) dépourvu de marque de temps.

8. Le terme «tamophorique» a été proposé par Tournadre (à paraître). Dans notre étude détaillée du phénomène (François 2003: 45-75), nous parlions de *noms aspectualisés*, raccourci pour *noms modo-aspecto-temporalisés*.

### 1.2.2 Mettre en perspective des instants ou des mondes

Les noms se combinent donc aux marques (T)AM si, et seulement si, la prédication nominale met en jeu des opérations de nature aspectuelle et/ou modale, distinctes de la simple assertion réalisée à valeur stative (type *X est/était... Y*).

Une prédication nominale sera aspectuellement marquée si elle présente le prédicat comme doté d'une structure temporelle interne, au sens où la propriété *p* se trouverait vérifiée sur une certaine portion de temps, et invalidée ailleurs. Ce type d'implication sémantique se trouve illustré par l'emploi de l'Accompli *may*, déjà rencontré en (1):

- (14) *Imam mino* <may tēytēybē>  
 père mon ACP guérisseur  
 'Ça y est, mon père est (~ était...) devenu médecin.' [aspect ACCOMPLI]

Alors qu'un prédicat nominal direct, de type (4), consistait à simplement valider une propriété *p* dans une situation de référence, sans rien dire des autres instants (glose *X = médecin*), on voit que le recours à une marque aspectuelle place ce même prédicat en perspective temporelle: en (14), on contraste une phase *p* (*X est médecin à l'instant de référence*) avec son complémentaire *p'* (*X n'était pas encore médecin*), y compris avec toutes les nuances sémantiques qui sont propres à l'opération aspectuelle choisie — l'Accompli impliquant normalement (contrairement au Parfait, par exemple) que la propriété *p* a été atteinte après avoir été visée, d'où la glose 'ça y est'.

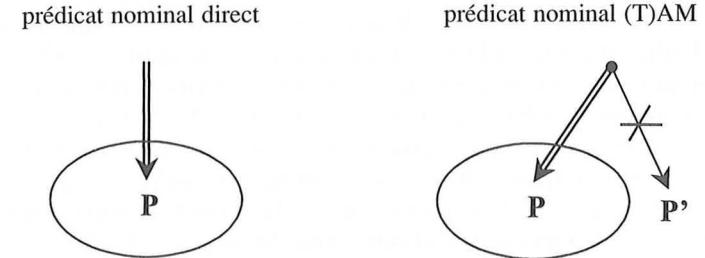
On retrouve des mécanismes assez semblables lorsque le morphème comporte une dimension modale, comme c'est le cas du Prioratif (*ni-... bah en*):

- (15) *Imam mino* <ni-tēytēybē bah> en!  
 père mon AO-guérisseur PRIO<sub>1</sub> PRIO<sub>2</sub>  
 'Attendons d'abord que mon père devienne médecin.' [aspect/mode PRIORITIF]

Ici aussi, la propriété *p* se trouve prise dans un contraste qualitatif entre une phase *p*, qui valide le prédicat (*X sera/serait médecin*), et une autre phase *p'* qui implique son contraire (*à l'heure actuelle, X n'est pas [encore] médecin*). Si l'on cherche à la définir, la différence entre valeurs aspectuelles et modales porte sur la nature des phases *p/p'* qui se trouvent posées en contraste: une opération aspectuelle — de type (14) — oppose entre elles deux phases temporelles à l'intérieur du même monde de référence (*X n'était pas Y, puis il l'est devenu*), alors qu'une opération modale — de type (15) — oppose plutôt deux mondes entre eux, l'un réel (*X n'est pas encore Y*), l'autre visé (*j'attends/je souhaite que X devienne Y*).

On peut résumer ces deux mécanismes en disant qu'en mwotlap, les noms se trouvent marqués en TAM à chaque fois que la propriété *p* se trouve «mise en perspective», et s'oppose à son complémentaire *p'* soit dans le temps, soit entre les mondes possibles. On peut représenter ce type de mécanisme cognitif par la *Figure 1*.

Figure 1 — Deux façons de prédiquer une propriété nominale



### 1.2.3 Propriétés grammaticales et économie du lexique

Au passage, on notera que dans les deux exemples (14) et (15), la traduction fait appel au verbe français *devenir*, lequel n'a pas d'équivalent en mwotlap. Et en effet, de même que la prédictivité de la catégorie NOM rend superflu l'usage d'une copule stative de type *être*, de même, la compatibilité totale des noms avec les marques aspectuelles et modales permet à la langue de se passer d'une copule dynamique / transformative, de type *devenir*: le sémantisme de cette dernière se trouve impliqué par le marquage grammatical en (T)AM. Les mêmes remarques s'imposent d'ailleurs du côté des adjectifs: dans la mesure où ceux-ci sont compatibles avec les mêmes marques aspectuelles ou modales que les verbes, la langue n'a besoin d'aucun autre outil grammatical pour distinguer la prédication d'état (*être P*) au Statif [ex. (16)], de l'évocation d'un changement d'état (*devenir P*):

- (16) *Ne-telefon* <ne-het>  
 ART-téléphone STA-mauvais  
 'Le téléphone est de mauvaise qualité / en mauvais état.' [aspect STATIF]
- (16') *Ne-telefon* <me-het>  
 ART-téléphone PFT-mauvais  
 'Le téléphone est tombé en panne.' [aspect PARFAIT]
- (16'') *Ne-telefon* <tiple het>  
 ART-téléphone ÉVIT mauvais  
 'Il ne faudrait pas que le téléphone tombe en panne.' [mode ÉVITATIF]

Qu'il s'agisse des noms ou des adjectifs, le mécanisme grammatical que nous décrivons — qu'on pourrait appeler «aspecto-modalisabilité», ou encore «tamorphoricité» (Tournadre) — présente des implications remarquables sur la structuration même du lexique. Outre l'absence des copules *être* ou *devenir* que nous venons de signaler, ce sont en fait tous les prédictats de transformation d'état qui se trouvent inclus dans l'expression de l'état stable correspondant. Ainsi, alors que la plupart des langues européennes distingueront l'adjectif (*être*) *rouge* du verbe *rougir*, ces deux valeurs distinctes — l'une stative, l'autre dynamique — seront encodées

identiquement en mwotlap, par l'adjectif *lawlaw* '(être/devenir) rouge'. De même, *het* traduira aussi bien '(être) mauvais' que 'se détériorer, tomber en panne...'; *wē* signifie à la fois 'bon, en bonne santé' et 's'améliorer, guérir...'; l'adjectif *leg* est à la fois '(être) marié' et 'se marier', etc<sup>9</sup>.

Enfin, pour revenir au domaine des noms, il suffira d'aspecto-modaliser le nom *tamayge* 'vieillard' pour traduire notre verbe 'vieillir', ou le nom *lōmgep* 'jeune garçon' pour signifier 'rajeunir' — sachant que, dans tous les cas, on aura toujours affaire à des noms, sans qu'il soit nécessaire de supposer un processus de dérivation en verbes. On voit donc comment la compatibilité des noms et des adjectifs avec les marques TAM, bien davantage qu'une simple propriété morphologique parmi d'autres, constitue en réalité un rouage central à l'économie de toute la langue.

#### 1.2.4 L'éventail sémantique des possibles

Nous venons d'examiner les principes sémantiques généraux de la combinaison des noms avec les TAM. Pourtant, s'il est vrai que ces mécanismes concernent virtuellement l'ensemble de la catégorie des noms, il n'empêche que d'un point de vue statistique, ces tournures demeurent minoritaires dans le discours: ainsi, sur un corpus de 75 000 mots, on ne trouve guère plus d'une quarantaine d'exemples de noms aspecto-modalisés — chiffre qu'il faut comparer avec les centaines de prédictats équatifs ou inclusifs, construits directement (§1.1.2).

Cette dissymétrie s'explique probablement par la propension fondamentale des noms à exprimer des propriétés aspectuellement stables: contrairement aux adjectifs, et surtout aux verbes, qui désignent le plus souvent des propriétés transitoires — et donc prédisposées à l'aspectualisation — les noms servent normalement à désigner leur référent au travers de son essence, laquelle est typiquement douée de permanence<sup>10</sup>. Ainsi, même si l'incompatibilité des noms avec le codage TAM, telle qu'elle est parfois postulée *a priori* dans la littérature, se trouve contredite par les faits du mwotlap — ainsi que d'autres langues (Nordlinger et Sadler 2000) —, elle n'en reflète pas moins une tendance lourde du point de vue typologique; tendance qui prend la forme d'une agrammaticalité absolue dans la plupart des langues du monde, et d'une simple infériorité statistique en mwotlap.

Or, c'est précisément parce que l'aspecto-modalisation des noms constitue un paradoxe du point de vue fonctionnel et typologique, qu'il peut être utile de passer en revue, brièvement, les domaines sémantiques où elle s'applique plus particulièrement. Par exemple, un nom comme *vētan* 'terrain,

9. Cette labilité fondamentale des noms et des adjectifs va de pair avec une originalité dans le domaine de l'actionalité verbale: tous les lexèmes statifs du mwotlap encodent également le prédictat dynamique correspondant, ex. *mitiy* 'dormir / s'endormir', *hey* 'porter / enfiler (un habit)'... Nous avons décrit ce phénomène sous le nom de Gabarit standard de Procès (François 2001 *a*; 2003: 97-104, 346-363).

10. Cf. Lemaréchal (1989: 33): «Les *noms* expriment des caractéristiques définitoires, les *adjectifs* des caractéristiques stables non définitoires, et les *verbes* des caractéristiques dont la validité est limitée à un procès (...), sinon à une énonciation».

sol' n'est pas attesté dans ce type de structures, parce qu'il est peu probable qu'advienne un contexte où l'on puisse dire d'un référent X qu'il est ??*devenu le sol...*; de même, la plupart des noms abstraits (ex. *na-mgaysēn* 'la compassion', *na-ñya* 'le rire', *nē-dēmdēm* 'idée, pensée'...) ne se trouvent jamais dans ce contexte. Inversement, certains domaines sémantiques rendent ces processus d'aspectualisation à la fois naturels et fréquents, comme nous allons le voir.

D'une manière générale, il est banal de tamophriser les noms qui renvoient à une essence instable dans le temps; c'est le cas, par exemple, de tous les lexèmes qui désignent une phase de croissance naturelle. Ce principe concerne par exemple les phases d'âge chez les humains: *nētñey* 'enfant' (≈2 à 10 ans environ); *yañfala* 'jeune', *lōmgep* 'jeune garçon', *ñalñmal* 'jeune fille' (≈10-18); *et liwo* 'grande personne', *tañan* 'homme', *lōqōvēn* 'femme' (≈18-60+); *tamayge* 'vieillard', *magtō* 'vieille femme' (≈60+), etc.

- (17) *Kē* ⟨n-et liwo⟩.  
3SG ART-personne grand  
'C'est un adulte.' [prédicat nominal direct]
- (17') *Kē* ⟨ni-et liwo galsi bah⟩ en!  
3SG AO-personne grand parfaitement PRIO<sub>1</sub> PRIO<sub>2</sub>  
'Attendons d'abord qu'il soit tout à fait (devenu) adulte!' [prédicat nominal TAM]
- (18) *Bōbō mino* ⟨mal qētēg magtō⟩.  
aïeul(e) mon ACP commencer vieille.femme  
'Ma grand-mère est déjà âgée.' [...a déjà commencé à (devenir) vieille-femme]

De façon comparable, la croissance des animaux ou des plantes, pour peu qu'elle implique des lexèmes spécialisés pour chaque phase de croissance (cf. fr. *têtard* vs. *grenouille*), donne souvent lieu à des noms marqués en TAM:

- (19) *Kē* ⟨so ni-proprok⟩ ēgēn!  
3SG PRSP AO-grenouille:DUP maintenant  
'Elle va bientôt (devenir) une grenouille.' (< angl. *frog*)
- (20) *Kē* ⟨ni-mañkē van i mañkē⟩ en, tō kē ⟨ni-et⟩.  
3SG AO-singe DIR DUR singe TOP alors 3SG AO-homme  
'(L'homme) a d'abord été longtemps singe, avant de devenir homme.'  
[lit. 'Il singea, il singea, puis il homma.']
- (21) *Nē-tqē* ⟨so ni-maltow⟩ en, tō so õñ ēgēn.  
ART-champ PRSP AO-brousse TOP alors PRSP défricher maintenant  
'Lorsque le champ (devient) brousse/jachère, il faut le défricher.'

Il peut s'agir d'un phénomène naturel, à condition qu'il soit limité dans le temps:

- (22) *Kē* ⟨ni-lo van i lo⟩ en, tō kē ⟨ni-sñal ganwōn⟩.  
3SG AO-soleil DIR DUR soleil TOP alors 3SG AO-pluie soudain  
'Après avoir longtemps «soleillé», il s'est mis soudain à pleuvoir.'

Tous les artefacts sont compatibles avec l'aspectualisation, dans la mesure où ils peuvent être placés dans la perspective de leur fabrication. Ainsi, en face du simple prédicat nominal *ni-siok* /ART-pirogue/ ‘c'est une pirogue’ [sur le modèle de (5)], on peut envisager le moment où l'on passe précisément de l'arbre à la pirogue, et donc de *p'* à *p*:

- (23) ⟨Mal siok⟩ *nen, si tateh qete?*  
 ACP pirogue cela ou NÉG (pas. encore)  
 ‘C'est déjà (devenu) une pirogue, ou pas encore?’

Il est un autre domaine sémantique où les noms réfèrent par excellence à des essences transitoires: c'est celui du statut social, et notamment du métier.

- (24) *Tō kē* ⟨ni-yogyogveg nōnōm⟩.  
 alors 3SG AO-serviteur ton  
 ‘C'est alors qu'il devenait ton serviteur.’
- (25) *Nok* ⟨so tēytēybē ne gatgat se⟩!  
 1SG PRSP AO:guérisseur de langue aussi  
 ‘Moi aussi j'aimerais bien (être/devenir) linguiste!’
- (26) *No ne-myōs so dōyō* ⟨so bulsal⟩.  
 1SG STA-vouloir que 1IN:DU PRSP AO:ami  
 ‘J'aimerais que nous devenions amis / que nous sortions ensemble.’

Il existe d'autres cas de figure encore, que nous ne détaillerons pas tous ici (cf. François 2003: 53-67).

Pour terminer ce survol des noms marqués en TAM, il faudrait classer à part certains énoncés, dans lesquels ce n'est pas le sujet qui se transforme au cours du temps, mais plutôt le parcours perceptif de l'observateur. Les contraintes sémantiques sur le sujet sont par conséquent distinctes des cas précédents:

- (27) ⟨May Lahlap⟩.  
 ACP (village)  
 ‘Ça y est, c'est déjà Lahlap (= nous y sommes arrivés/\*c'est devenu Lahlap).’

Les restrictions sémantiques énumérées plus haut perdent également de leur pertinence lorsque la valeur de la proposition nominale est purement modale (ici, le Contrefactuel):

- (28) *Nēk* ⟨te-lqōvēn tō⟩ *en, togtō nok leg mi nēk!*  
 2SG CF<sub>1</sub>-femme CF<sub>2</sub> TOP alors:CF 1SG AO:marié avec 2SG  
 [plaisanterie] ‘Si tu étais une femme, je me marierais avec toi!’

Dans la mesure où l'énoncé s'adresse à un homme, le nom *lōqōvēn* ne reçoit pas ici sa signification de phase d'âge (‘femme adulte’, opp. ‘jeune fille’), mais de genre (‘femme’ opp. ‘homme’): la notion d'essence transitoire n'est donc pas en jeu ici.

### 1.3 Une liberté sous contraintes

Le mwotlap, nous l'avons vu, attribue donc à plusieurs catégories, dont celle des noms et des adjectifs, des propriétés qui leur sont interdites dans les langues européennes: d'une part, la prédicativité directe; d'autre part, la totale compatibilité avec les opérations aspectuelles et modales. Cette configuration, typologiquement assez originale, peut donner l'impression d'un système beaucoup moins contraint que les autres langues, dans lequel tout lexème pourrait constituer librement un prédicat, et recevoir des marques TAM sans aucune restriction. Pourtant, le phénomène connaît certaines limites, que nous présenterons brièvement — ne serait-ce que pour mieux comparer le mwotlap avec d'autres langues d'Océanie, dans lesquelles les restrictions sont différentes.

#### 1.3.1 Des restrictions selon les catégories

Dans la section 1.1, nous avons passé en revue les différents types de prédicats du mwotlap: d'un côté, des prédicats TAM, de l'autre, des prédicats directs de divers types. Dans un second temps (§1.2), nous avons vu que si les verbes et les adjectifs ne rentraient que dans la première de ces catégories, la catégorie des noms, en revanche, était théoriquement compatible avec les deux structures syntaxiques, en fonction du sémantisme de la prédication (cf. *Figure 1*).

Une première hypothèse consisterait à hiérarchiser entre elles les deux types de structures, selon une formule du type { PRÉDICAT TAM < PRÉDICAT DIRECT }. Cela signifierait qu'étant donné une catégorie syntaxique, si elle est compatible avec les prédicats directs, alors elle l'est également avec les prédicats TAM (ex. les noms), la symétrique n'étant pas vraie (ex. les adjectifs). Et en effet, cette hypothèse semble confirmée par la classe des numéraux, qui sont effectivement compatibles avec les deux structures, moyennant le même type de contraintes sémantiques que nous avons examinées pour les noms. Alors que la structure directe de type (10) attribuait simplement au sujet un nombre (*le nombre de X s'élève à N*), la combinaison à des marques TAM consiste à établir un contraste entre *p* et *p'*, qu'il s'agisse de phases temporelles ou de mondes possibles:

- (29) *Talōw n-ête nonon* ⟨ni-vētēl⟩.  
 demain ART-année sa AO-trois  
 ‘Elle aura trois ans demain.’ [iitt. ses années (deviendront) trois]
- (30) *Velqōn, na-kat* ⟨so ni-levete⟩ *le-mnē.*  
 toujours ART-cartes PRSP AO-six dans-main:2SG  
 ‘Tu dois toujours avoir six cartes dans la main.’  
 [iitt. Que les cartes «sixent» dans ta main.]

Il en va de même, quoique de façon limitée, pour les pseudo-adjectifs, les possessifs et les prédicats existentiels (François 2003: 72-75). Pourtant, un contre-exemple majeur est constitué par les prédicats locatifs, lesquels sont

absolument incompatibles avec toutes marques TAM. Si l'on voulait pourvoir l'énoncé (6') d'une valeur aspectuelle ou modale, on serait obligé de recourir à la médiation d'un verbe, ex. *dēñ* ‘atteindre’:

- (6") *Bulsal mino* <may dēñ> *Iñglan.* [\*...may Iñglan.]  
 ami mon ACP atteindre Angleterre  
 ‘Ça y est, mon ami est (litt. est arrivé) en Angleterre.’

De cette observation, il résulte que la prédicativité d'une part, et la sensibilité au temps-aspect-mode d'autre part, constituent deux paramètres indépendants, qui se trouvent distribués différemment selon les catégories syntaxiques de la langue (*Tableau 1*).

**Tableau 1** — Compatibilité des principales catégories syntaxiques avec la prédicativité et le marquage TAM

	Nom	Adjectif	Attribut	Verbe	Exist <sup>tif</sup>	Locatif	Numéral
Prédicat direct	+	-	+	-	+	+	+
Prédicat TAM	+	+	(+)	+	(+)	-	(+)

### 1.3.2 Une limite sémantique

Nous terminerons en évoquant la difficulté d'analyser certains énoncés ambigus; ceci nous conduira à observer les limites de la prédication nominale. Jusqu'à présent, nous avons adopté une position ferme: dans les énoncés où un lexème nominal se trouve combiné à une marque TAM, celui-ci demeure un nom, sans qu'il soit ni nécessaire, ni légitime, de considérer qu'il a été dérivé en verbe. Ce type d'analyse reposeraient uniquement sur le préjugé que le temps-aspect-mode constituerait une propriété intrinsèquement verbale, préjugé que précisément une langue comme le mwotlap permet d'infirmer.

Un argument, parmi d'autres, consiste à observer que le marquage en TAM n'empêche pas la tête prédicative de garder une syntaxe nominale, comme la qualification par un adjectif en (17'), par un syntagme déterminatif en (25), ou par un possessif en (24). Par conséquent, le lexème *tita* ‘mère’ est bien un nom en (31) et en (31'):

- (31) *Kē* <*tita mino*>. 3SG mère ma  
 ‘C'est ma mère.’ [NOM prédicat équatif]
- (31') *Kē* <*ni-tita mino*> *ēgēn*. 3SG AO-mère ma maintenant  
 ‘Du coup, elle devient ma mère.’ [NOM prédicat TAM]

Mais la réponse est moins évidente dans le cas de certaines tournures, dans lesquelles la syntaxe du lexème devient typiquement verbale (prédicat

à deux arguments de type verbe transitif, incompatibilité avec les déterminants nominaux):

- (31'') *Kē* <*ni-tita*> no *ēgēn*. 3SG AO-avoir.qqn.pour.mère 1SG maintenant  
 ‘Du coup, il/elle me traite comme sa mère.’ [nom converti en VERBE]

Plusieurs arguments suggèrent ici qu'on n'a plus affaire à un nom, mais à un verbe dénominal (obtenu par dérivation zéro ou conversion). Premièrement, ce type d'ambiguïtés ne concerne qu'une poignée de noms en mwotlap — en particulier les noms de parenté — et ne peut donc pas être présenté comme une caractéristique globale de la partie du discours NOM. Deuxièmement, la syntaxe de l'énoncé (31'') est de type verbal et non nominal. Troisièmement, la tête prédicative change à la fois de signification lexicale et d'orientation diathétique, puisqu'en (31') *tita* signifie [X] *être/devenir la mère de [Y]*, alors qu'en (31'') le même mot reçoit une interprétation opposée [Y] *avoir pour mère [X]* (*considérer qqn comme sa mère*).

Ces arguments nous conduisent à voir en (31'') un énoncé verbal, et nous permettent de définir les limites du phénomène que nous avions décrit en §1.2: certes, les noms peuvent former des prédicats directs et même porter des marques TAM sans cesser d'être des noms à part entière; mais ce, à la condition de préserver une syntaxe nominale, et plus précisément de préserver la même diathèse que si le lexème formait un prédicat nominal direct (*X être P*  $\Leftrightarrow$  *X devenir P*). Voilà qui prouve que la mince cloison qui sépare les noms des verbes ne doit pas être abattue trop vite. Pourtant, s'il est vrai que cette analyse semble s'imposer dans une langue comme le mwotlap, nous allons voir bientôt que des énoncés similaires peuvent recevoir une interprétation différente dans d'autres langues d'Océanie.

## 2 La prédication non-verbale dans d'autres langues d'Océanie

Sur la question des prédicats non-verbaux, le mwotlap, que nous venons d'observer en détails, est globalement assez représentatif des langues de la famille océanienne. Pourtant, à y regarder de plus près, on constate une relative diversité parmi les configurations syntaxiques attestées dans ce vaste ensemble de langues. C'est ce que nous allons voir brièvement, en observant quelques langues de Mélanésie et de Polynésie.

### 2.1 Les paramètres syntaxiques

#### 2.1.1 La copule

En général, les langues océaniennes, comme le mwotlap, sont dépourvues de copule, aussi bien au présent qu'aux autres temps/aspects — avec pour corollaire, la compatibilité des noms avec les marques TAM. Pourtant, un petit nombre de langues parlées au centre du Vanuatu, recourent à un

verbe *être* dans les mêmes conditions que le russe ou l'arabe. C'est par exemple le cas du paama (Lynch *et al.* 2002: 49):

- (PA.1) Mail ro vi -tei asuv.  
 (nom) 3SG:REA:NÉG<sub>1</sub>- être -NÉG<sub>2</sub> chef  
 'Mail n'est pas chef.'

Dans certaines langues polynésiennes, les prédictats nominaux équatifs *stricto sensu* font normalement appel à une particule \*ko parfois glosée 'présentatif', ou 'identificateur'. Ce morphème, qui peut être lui-même marqué en temps, pourrait s'apparenter à une sorte de copule invariable, dans la mesure où il constitue, en un sens, le support syntaxique de la prédication nominale. Voici un exemple en wallisien (Moyse-Faurie, à paraître):

- (WA.1) ⟨Ne'e ko te 'aliki⟩.  
 PASSÉ PRÉS ART chef  
 'C'était le chef.'

La situation est en réalité plus complexe, car le même 'présentatif' / 'identificateur' remplit également d'autres fonctions, notamment celles de topicalisateur ou d'article pour les noms propres (Lazard et Peltzer 2000). Ceci dit, même si l'on ne saurait le réduire à une simple copule, ce type de morphème incite à nuancer la notion d'omniprédictativité dans le cas précis des langues polynésiennes (Gilbert Lazard, comm. pers.).

### 2.1.2 Prédicats directs ou prédicats TAM?

D'une façon générale, l'éventail des prédicats, tel que nous l'avons présenté en 1.1 pour le mwotlap, se retrouve dans les autres langues de la famille: on rencontre ainsi des prédicats directs constitués de possessifs, de numéraux, d'existentiels, de locatifs, etc.

Parfois, les numéraux ne peuvent prédiquer que s'ils sont marqués en TAM, et se comportent donc comme une sous-classe de verbes: c'est le cas en araki, autre langue du Vanuatu (François 2002: 82). De la même façon, les structures syntaxiques dans lesquelles entrent noms et verbes se distinguent parfois moins nettement qu'en mwotlap. Ceci peut s'expliquer soit parce que des marques TAM segmentales sont requises dans les deux cas; soit parce qu'une des marques TAM a la forme Ø, ce qui rend opaque la différence entre «prédicats directs» et «prédicats TAM» — cf. le xârâcùù (Moyse-Faurie 1995: 145):

- (XA.1) Nâ (Ø) aaxa. / Nâ (Ø) pûxûrû.  
 1SG (PST) chef 1SG (PST) courir  
 'Je suis chef.' 'Je cours.'

Enfin, l'absence totale de temps grammatical en mwotlap, que nous avons observée autant dans les prédicats TAM que les prédicats directs (§1.2.1), constitue une originalité y compris au niveau de la famille océanienne: la plupart des autres langues différencient, par exemple, le présent du passé — cf. ex. (WA.1) ci-dessus.

### 2.2 Les paramètres sémantiques

Le dernier type important de différences entre langues océaniennes concerne l'interprétation sémantique de la prédication elle-même: car dire que deux langues autorisent les noms à la place de prédicat, n'implique pas nécessairement que la structure obtenue doive recevoir la même interprétation sémantique. Pour prendre l'exemple des locatifs en prédicat direct, le mwotlap { *Sujet + Locatif* } signifiera 'X se trouve en L', alors que la même structure en wallisien aura une valeur de provenance géographique:

- (WA.2) ⟨'E 'Uvea⟩ tefafine 'aia.  
 NON-PASSÉ Wallis ART femme ANAPH  
 'La femme en question est wallisienne [\*se trouve à Wallis].'

La variété des interprétations concerne surtout les prédicats nominaux. Certes, on retrouve souvent des prédicats à valeur équative ou inclusive (*X est un/le N*) comme en mwotlap — cf. (WA.1), (XA.1). Cependant, on trouve également des noms prédicats dotés d'une application «situationnelle» plutôt que «subjectale» (Launey 1994: 35): ainsi, en xârâcùù, *Daa* /jour/ signifie à lui seul 'il fait jour' (plutôt que \*c'est un jour), et de même, la tournure à nom dépendant *Purè-mwâ* /le.vide-maison/ devra se traduire 'la maison est vide' (Moyse-Faurie 1995: 143 sq.).

Les langues polynésiennes rendent la prédication nominale encore plus polysémique. Ainsi, alors que le prédicat 'pirogue' en mwotlap ne pourrait recevoir comme sujet que l'embarcation elle-même (cf. ex. 23), en futunien le même lexème peut avoir pour sujet l'agent qui utilise la pirogue, en vertu donc d'une interprétation non pas équative, mais instrumentale:

- (FU.1) ⟨'E vaka⟩ a Petelo i le vaka fea?  
 NON-PASSÉ bateau ABS (nom) OBL ART bateau quel  
 'Sur quel bateau est Petelo?' (litt. Petelo [fait du] bateau sur quel bateau?)

D'autres valeurs encore sont attestées, du type *X a la propriété de N* (ex. 'le chemin cailloute'), ou *X est le moment de N* ('j'ai dimanché avec elle'), etc.; la diversité des cas de figure se trouve abordée en détails dans Moyse-Faurie (à paraître), ainsi que dans Lazard et Peltzer (2000) pour le tahitien. Le point important est de voir que les analyses syntaxiques que nous avions proposées pour le mwotlap, et qui consistaient à parler de conversion de nom en verbe dès lors que la diathèse du lexème se trouvait altérée (§1.3.2), ne s'imposent peut-être pas avec la même nécessité dans des langues où la prédication nominale elle-même se trouve investie de plusieurs significations distinctes: c'est ainsi que, selon l'analyse de Moyse-Faurie, l'énoncé (FU.1) ci-dessus demeure un prédicat nominal.

### 3 Conclusion

Au regard de la prédication non-verbale, les langues océaniennes frappent d'abord par leur unité: pour la plupart, elles ne possèdent pas l'équivalent de

notre verbe *être*, et reportent directement sur le terme attribut (adjectif, nom, possessif, numéral, locatif, proposition...) le trait de prédicativité, voire les opérations mêmes de temps-aspect-mode. Pourtant, à y regarder de plus près, cette famille connaît en son sein une certaine diversité, aussi bien syntaxique que sémantique, au point que chaque configuration particulière impose de nuancer les analyses proposées pour la langue voisine: tel système fera des adjectifs ou des numéraux une sous-classe de verbes; tel autre système placera la frontière entre prédicats nominaux et verbaux à tel endroit, lequel peut être différent dans d'autres langues pourtant apparentées... Voilà qui devrait inciter à poursuivre les recherches de terrain pour décrire avec précision les langues de cette famille, dont la plupart sont encore trop mal connues.

#### 4 Abréviations

ABS	absolutif	EXIST	prédicat existentiel
ACP	aspect Accompli	OBL	oblique
ANAPH	anaphorique	PRÉS	présentatif / identificateur
AO	aspect Aoriste	PRIOR	aspect Prioritif
ART	article	PRSP	aspect Prospectif
CF	mode Contrefactuel	PST	présent
COP	copule	RÉA	Réalis
DIR	directionnel	STA	aspect Statif
DUP	réduplication	TAM	Temps-Aspect-Mode
DUR	duratif	TOP	topicalisateur

#### Références

- FRANÇOIS A. (2001a), «Gabarit de procès et opérations aspectuelles en motlav (Océanie)». *Actances* 11, 145-175.
- FRANÇOIS A. (2001b), «Contraintes de structures et liberté dans l'organisation du discours. Une description du mwotlap, langue océanienne du Vanuatu», thèse de Doctorat, Université Paris-IV Sorbonne, Paris, 3 vol., 1078 pages.
- FRANÇOIS A. (2002), *Araki. A disappearing language of Vanuatu*, Pacific Linguistics 522, Canberra, Australian National University.
- FRANÇOIS A. (2003), *La sémantique du prédicat en mwotlap (Vanuatu)*, Collection Linguistique de la Société de Linguistique de Paris, 84, Paris, Louvain, Peeters.
- LAUNAY M. (1994), *Une grammaire omniprédictive: Essai sur la morphosyntaxe du nahuatl classique*, Sciences du Langage, Paris, CNRS.
- LAZARD G. et PELETZER L. (2000), *Structure de la langue tahitienne*, Langues et Cultures du Pacifique, 15, Paris, Louvain, Peeters.
- LEMARÉCHAL A. (1989), *Les parties du discours, Syntaxe et sémantique*, Linguistique Nouvelle, Paris, PUF.
- LYNCH J.; ROSS M. et CROWLEY T. (2002), *The Oceanic languages*, Richmond, Curzon Press.

- MOYSE-FAURIE C. (1995), *Le xârâcùù : Langue de Thio-Canala (Nouvelle-Calédonie)*, Langues et cultures du Pacifique, 10, Paris, Louvain, Peeters.
- MOYSE-FAURIE C. (1997), *Grammaire du futunien*, Collection Université, Nouméa, Centre de Documentation Pédagogique.
- MOYSE-FAURIE C. (à paraître), Problèmes de catégorisation syntaxique dans les langues polynésiennes, in *Linguistique typologique: Actes du Troisième Colloque de Typologie* (2002), Lille, Septent.
- NORDLINGER R. et SADLER L. (2000), Tense as a nominal category, Communication à la conférence *Lexical Functional Grammar 2000*, Berkeley, Juillet 2000.
- TESNIÈRE L. (1953), *Esquisse d'une syntaxe structurale*, Paris, Klincksieck.
- TOURNADRE N. (à paraître), «Typologie des aspects verbaux et intégration à une théorie du TAM», *Bulletin de la Société de Linguistique de Paris*.

Alexandre FRANÇOIS  
LACITO-CNRS  
7 rue Guy Môquet  
94801 Villejuif

Alexandre.Francois@vjf.cnrs.fr

# Serial Verb Constructions

---

*A Cross-Linguistic Typology*

edited by

ALEXANDRA Y. AIKHENVALD  
and

R. M. W. DIXON

Research Centre for Linguistic Typology,  
La Trobe University

10

---

## Serial Verb Constructions in Mwotlap

*Alexandre François*

Mwotlap is an Austronesian language of the Oceanic branch, spoken by about 1,800 speakers on Motalava, a small island of the Banks group, north of Vanuatu (François 2001, 2003). Contrary to many languages from the same area—for example, Paamese (Crowley 1987), Ambae (Hyslop 2001), Araki (François 2002)—Mwotlap has almost no traces of serialization between verb phrases (of the type *I'll PUSH you you'll FALL*).<sup>1</sup> The only productive case of verb serialization is of a different pattern, with two or more verbs chained together within a single verb phrase (of the type *I'll PUSH FALL you*). A typical example of this serializing construction can be heard in a famous love song:<sup>2</sup>

- (1) [lak tēy yoyoñ ēwe] no  
aor:dance hold be.quiet be.fine me  
'Just dance with me calmly'

Simple though it may be, a sentence like (1) raises a number of issues. What are the relations, both syntactic and semantic, between the four verbs? How do they combine their lexical semantics, their tense–aspect–mood values, their argument structures? How do such constructions compare with other serializing patterns cross-linguistically?

After situating Mwotlap SVCs in their syntactic context (§1), we will describe their structural properties (§2), paying special attention to the sharing of arguments (§3). We will then propose a functional classification of these serial verbs (§4), and end our reflection with a note on multiverb serialization (§5).

<sup>1</sup> To be precise, Mwotlap does possess a type of pauseless parataxis that is reminiscent of the so-called 'core-layer serialization', such as *Give me some water I drink it* (François 2003: 188). However, it appears that this construction is limited to a specific TAM context (the second verb must be inflected as aorist) and to a single semantic value (the expression of purpose). Consequently, this structure is probably better described as a case of paratactic subordination involving two distinct clauses, rather than a standard case of verb serialization.

<sup>2</sup> The spelling conventions adopted for Mwotlap include the following: *e* = [ɛ]; *ē* = [ɪ]; *o* = [ɔ]; *ō* = [ʊ]; *g* = [ɣ]; *b* = [ʷb]; *d* = [ⁿd]; *q* = [kpʷ]; *m̄* = [ŋmʷ]; *n̄* = [ŋ].

## 1. Clause structure and verb serialization

### 1.1. A PRELIMINARY NOTE ON WORD CLASSES

The inventory of word classes in Mwotlap makes it possible to distinguish between verbs and adjectives. For example, verbs cannot modify a noun directly, whereas adjectives can: compare *nētñey gom<sub>[adj]</sub>* ('a sick child') and \**nētñey teñ<sub>[vb]</sub>* (\*'a cry child'). And yet, adjectives and verbs behave exactly the same way outside noun phrases. For example, both categories require TAM markers in order to form a valid predicate phrase, without any copula:

- (2) (a) inti-k me-teñ  
child-1sg PER-cry  
'My son is crying'  
(b) inti-k mo-gom  
child-1sg PER-sick  
'My son is sick'

Since the study of serial structures is not concerned with noun phrases, it will be legitimate, for our present purposes, to consider adjectives as forming a subclass of (intransitive, stative) verbs.

### 1.2. THE STRUCTURE OF THE CLAUSE

The standard order of constituents in Mwotlap is SV for intransitive and AVO for transitive clauses, which is typical of a nominative-accusative system. In the absence of any sort of case-marking, the syntactic function of the core arguments is indicated by their syntactic position. Verbs are either strictly intransitive or strictly transitive, a few being ambivalent (mainly of the S=A type). Mwotlap does not allow for double-object constructions.

Although TAM markers often consist of just a prefix, some of them are discontinuous, combining a prefix and a postclitic, like the Potential *te-*...*vēh* or the negator *et-*...*te*. This morphological property makes them a convenient tool to test the boundaries of the verb phrase, as they clearly show which constituents belong inside vs. outside the VP. For instance, in (3) below, the position of *vēh* allows us to distinguish between two distributional word classes, which in English would correspond to the single category of 'adverbs' (*again* and *tomorrow*):

- (3) kōmyō [TE-gen *lok se vēh*] na-madap *talōw* le-mtap  
2du POT<sub>1</sub>-eat back again POT<sub>2</sub> ART-pineapple tomorrow in-morning  
'You'll be able to eat pineapple once again tomorrow morning'

Reserving the term 'adverb' for those peripheral complements which always appear outside the VP (e.g. *talōw* 'tomorrow'), we propose the term 'adjunct' (Crowley 1982: 162) to designate those modifiers which belong inside the VP, and

appear immediately after its head (e.g. *lok* 'back' and *se* 'again'); we'll come back to this notion below.

We now have enough information to state the canonical structure of the sentence in Mwotlap:

{ subject [TAM<sub>1</sub> -HEAD adjuncts TAM<sub>2</sub>]<sub>VP</sub> object adverb/oblique }

Note that the object phrase is always external to the VP, unless we are dealing with an incorporated object (see §2.1).

### 1.3. THE NATURE OF ADJUNCTS AND THE LIMITS OF SVCS

Rather than a lexical category, the term 'adjunct' designates a syntactic position in the clause—that is, any word that appears within a predicate phrase, immediately following its head. Crucially, this position of VP-internal modifier can be lexified by more than one word class. First, Mwotlap possesses a category of 'pure adjuncts', which cannot appear anywhere else in the sentence other than that position. These were illustrated in (3), with *lok* and *se*.

But in some cases, the adjunct slot can also be filled by a noun:

- (4) Tigsas kē et-wot vu te, kē mo-wot et  
Jesus 3sg NEG<sub>1</sub>-be.born spirit NEG<sub>2</sub> 3sg PER-be.born person  
'Jesus was not born a spirit, he was born a man'

It is also common to find an adjective in the same position:

- (5) Imam ma-hag *qaqa ēwē l-ēm*  
Dad PER-sit stupid just in-house  
'Dad is just staying idle/idly at home'

And in many cases, the adjunct position is filled by a verb:

- (6) nitog hohole galgal!  
PROHIB talk:REDUP lie:REDUP  
'Stop lying!'

Following the discussion in §1.1, the description of serial verb constructions will be focused on sentences such as (5) as well as (6).

It is also worth mentioning cases in which the adjunct slot is filled by a word that was formerly a verb, but has now become a pure adjunct:

- (7) nok [tig day] bulsal mino  
1sg AOR:stand (expect) friend my  
'I'm (standing) waiting for my friend'

In a former stage of the language, when *day* could still behave as a verb in its own right, a sentence like (7) would have had to be described as a serial verb construction ('stand wait'), just like (6). But although it has retained certain features typical of verbs—such as a transitive argument structure—the lexeme

*day* is now restricted to this modifying position, as though it had become a sort of applicative clitic. As it no longer satisfies the definition of a verb—that is, compatibility with the position of head in a verb phrase—it is methodologically necessary to exclude such cases from our synchronic study of serial verbs. We will, however, come back to this issue in the conclusion.

## 2. Structural properties of Mwotlap SVCs

A serial verb construction can consist of two or more elements; the longest string attested is four verbs. We will start by examining the rules for ‘simple’ serial verbs ( $V_1 + V_2$ ), and will return to multiverb serialization in §5.

The formal properties of Mwotlap SVCs can be stated according to the typological criteria and terminology set out in the first chapter of this book. These are given in Table 1, and will be addressed separately in the following discussion.

### 2.1. CONTIGUITY

The two verbs forming an SVC must be strictly contiguous, that is, no element can intervene between them, whether this is an object or an oblique phrase. Even ‘pure adjuncts’, which are allowed in the VP, are not normally inserted between two serialized verbs. The only apparent exception to this rule is when the object of  $V_1$  is incorporated. In this case, the object  $O_1$  is suffixed to  $V_1$ , and therefore surfaces between the two verbs, as in:

- (8) nok [suwyeg-qen tēy] nu-sus  
 1sg AOR:cast-net hold ART-shoes  
 ‘I go net-fishing with my shoes on’

This construction—which is very rare anyway—is easily explained if one realizes that the first element in the SVC is not the transitive verb *suwyeg* ‘cast’, but an intransitive, compound verb of the form *suwyeg-qen* /cast-net/ ‘to net-fish’, with an incorporated object. A sentence like (8) is therefore no exception to the rule of strict contiguity between  $V_1$  and  $V_2$ .

TABLE 1. Main structural properties of Mwotlap SVCs

<i>Contiguity <math>V_1/V_2</math>?</i>	Strict contiguity
<i>Wordhood <math>V_1/V_2</math>?</i>	One prosodic word, two phonological words
<i>Symmetry <math>V_1/V_2</math>?</i>	Asymmetrical serial verbs (major + minor)
<i>Tense–aspect–mood, negation</i>	Identical + single-marking
<i>Pluractionality (reduplication)</i>	Non-identical + independent marking
<i>Argument sharing constraints</i>	Fusion rules (2 structures merging into 1)

### 2.2. ONE OR TWO WORDS?

The two parts of an SVC are so close to each other that it is sometimes tempting to analyse the string  $V_1 + V_2$  as an instance of verb compounding. This brings up the issue of wordhood in SVCs: are we dealing with one or two words?

At first sight, several arguments may advocate a one-word analysis. Indeed,  $V_1$  and  $V_2$  are not only strictly contiguous, but they also form a single prosodic unit: for example, *yow veteg* /jump leave/ will have only its final syllable stressed [jɔwβε‘tɛy] and no intonation break, like a single word. Semantically, the whole string  $V_1 + V_2$  often ends up being endowed with certain semantic features that can be assigned to neither of its components, as though it now formed a single unit: thus, *yow veteg* /jump leave<sub>TR</sub>/ means ‘escape from (someone, something)’, without implying any real ‘jumping’ event; *dēm veteg* /think leave<sub>TR</sub>/ is the usual way to translate ‘give up (something), forgive (someone)’. Yet, this is not sufficient evidence to conclude that we are dealing with a single word, as it is common for lexical units to consist of several words.

Another piece of evidence will ultimately prove that SVCs form distinct phonological words. Indeed, the phonotactic rules of Mwotlap make it possible to strictly identify the boundaries of the word. The only syllable pattern allowed in this language (François 2000) has the form CVC (with optional consonants), so that the phonotactic template of any word is #CVC|CVC...|CVC#. Consequently, clusters of two consonants are only allowed across syllable boundaries within a word, and never word-initially. There are two possible outcomes when a lexical root of the form C<sub>1</sub>C<sub>2</sub>V- has to be integrated in a sentence:

- if the root is in the position to begin a new syllable (typically after a word boundary), then the sequence #C<sub>1</sub>C<sub>2</sub>V- undergoes a vowel epenthesis, whereby a clone of V is inserted between the two consonants: for example /VTEG/ ‘leave’ → *nēk so veteg* ‘you should leave it’;
- conversely, if the same root is preceded by a vowel-final prefix, then the prefix + lexeme string forms a single word. The syllable boundary occurs between C<sub>1</sub> and C<sub>2</sub>, with no need for epenthesis: e.g. /VTEG/ → *nēk te-pteg* ‘you will leave it’.

In this pair of examples, the behaviour of the root /VTEG/ with regard to vowel epenthesis makes it clear when we are dealing with a single phonological word (form *te-pteg*: hence *te-* ‘Future’ is a prefix) or with two distinct words (form *so veteg*: hence *so* ‘Prospective’ is not a prefix). We can now apply the same phonological test to our serial verbs—provided the first verb ends in a vowel, and the second verb has an underlying CCV- root. If we consider the combination of  $V_1$  /hō/ ‘paddle, travel in canoe’ with  $V_2$  /VTEG/, the surface form we observe (‘paddle away’) is *hō veteg*, not \**hō-pteg*:  $V_1$  and  $V_2$  are thus separated by a word boundary, and cannot be said to form a single, compound word. As a conclusion, serial verbs in Mwotlap always remain distinct phonological words, whatever their degree of semantic or prosodic cohesion.

Finally, from the morphological point of view, examples (11a–b) below will show that root reduplication affects independently each element in a serial construction. This is also a strong argument in favour of the conclusion that Mwotlap SVCs fundamentally consist of separate words.

### 2.3. SHARING VERBAL CATEGORIES

If the serial verbs of Mwotlap were to be compared with other languages, they would probably stand at one end of the typological spectrum, that labelled ‘prototypical serial verbs’ in Chapter 1, and characterized by the highest degree of cohesion between its elements. Indeed, SVCs essentially behave the same as a single lexeme, with regard to almost all the semantic categories that may affect a verb phrase. Thus, all tense–aspect–mood markers are obligatorily shared by  $V_1$  and  $V_2$ , and they are marked only once:

- (9) kēy [to-yōnteg vēglal vēh] na-lñē  
 3pl POT<sub>1</sub>-hear know POT<sub>2</sub> ART-voice:2sg  
 ‘They might recognize your voice’

In (9), the Potential marker *te-vēh* appears once, and is shared by the two verbs; to use the terms of Chapter 1, Mwotlap SVCs are characterized by ‘single marking’ of TAM.

The same observation is true for negative markers, which in this language belong to the TAM paradigm rather than combine with it. Elements of an SVC cannot be negated separately, even if, semantically speaking, only one verb (here  $V_2$  *maymay*) falls under the scope of the negation:

- (10) kōyō may leg, ba [et-leg maymay qete]  
 3du COMPL married but NEG<sub>1</sub>-married strong NEG<sub>2</sub>:COMPL  
 ‘They’re already married, but not fully married yet’

There seems to be only one semantic category that is assigned independently to each member of an SVC: this is pluractionality, which is morphologically coded by root reduplication (François 2004b). In the next example, one may contrast different combinations, according to whether  $V_1$  refers to one ‘stoning’ event (*yim*) or to several (*yimyim*); and whether  $V_2$  refers to one death (*mat*) or to several (*matmat*):

- (11) (a) no [mi-yim matmat] ne-men  
 1sg PER-stone die:REDUP ART-bird(s)  
 ‘I stoned the birds (once) and killed them’  
 (b) kem [mi-yimyim mat] ne-men  
 1exc:pl PER-stone:REDUP die ART-bird(s)  
 ‘We stoned the bird(s) and killed it/them outright’

Finally, another important issue deals with the sharing of argument structures in serial verbs; this will be the topic of §3.

### 2.4. SYNTACTIC ASYMMETRIES OF $V_1/V_2$

The properties of Mwotlap SVCs reviewed thus far tend to suggest we are dealing with two verbs  $V_1$  and  $V_2$  placed on the same syntactic level, so that it might be tempting to talk about a non-hierarchized, multiheaded structure. In fact, several arguments show that  $V_1$  and  $V_2$  have a distinct status, and that their combination remains asymmetrical.

First of all,  $V_1$  and  $V_2$  do not have the same inventory. If all verbs are attested in the  $V_1$  slot, it is not true they can all function as  $V_{\pm 2}$ : such common verbs as *van* ‘go’, *vap* ‘say’, *yōnteg* ‘feel’, or *dēm* ‘think’ are attested only as  $V_1$ , and never as  $V_2$ . In other words, the inventory of verbs that can be serialized, however numerous, appears to constitute a (semi) closed list; the choice of  $V_2$  is clearly not as free as it seems at first.

Among other elements that betray an asymmetry between  $V_1$  and  $V_2$ , a handful of verbs show morphological differences according to their position. The verb ‘know’ has the form *ēglal* when used alone or as a first verb in a series, but becomes *vēglal* in the position of  $V_2$ , as in (9). The verb *sok* is reduplicated as *soksok* when used alone or as  $V_1$ , but as *sosok* when  $V_2$ ; similarly, *tēy* ‘hold’ reduplicates as *tētytēy* if  $V_1$ , but *tētēy* if  $V_2$ , and so on. These ‘SVC specific forms’ are seen only in this adjunct position.

The difference between  $V_1$  and  $V_2$  is even more striking if we begin to consider semantics. Quite often, a verb lexeme will keep its proper meaning when it is used as  $V_1$ , but will receive a more abstract or figurative interpretation when used as a verb modifier. To take just a couple of examples, *tēy* normally means ‘hold in one’s hands’ when in head position, but has a broader comitative meaning (‘be or act with someone/something’) when it acts as a verb modifier, as in (1) and (8). Similarly, *v(e)teg* as  $V_1$  means ‘lay (something) down, take leave of (someone)’; but as  $V_2$ , its more abstract meaning ‘away from (something/someone)’ allows for figurative uses such as ‘leave, forget, forgive, surpass’ (see §2.2).

All these arguments tend to confirm that the SVCs of Mwotlap, despite apparently forming a balanced string of two verbs  $V_1 + V_2$ , illustrate in fact what the typological chapter of this book called ‘asymmetrical serial verbs’, whereby a ‘minor verb’ from a closed class (adjunct  $V_2$ ) is being serialized to a ‘major verb’ from an open class (head  $V_1$ ).

## 3. Sharing arguments in Mwotlap SVCs

### 3.1. BASIC PRINCIPLES

In §2.3, we saw that the SVCs of Mwotlap are characterized by a strong internal cohesion, so that they necessarily share the same value in tense–aspect–mood or in polarity. The issue of argument-sharing is much more complex, and deserves to be examined in detail. It will appear that Mwotlap challenges certain typological statements in this regard.

The principles of our analysis are as follows. While each member of an SVC is lexically endowed with its own underlying argument structure, when serialized they behave exactly like a single verb: in particular, the SVC can have no more than one subject and one object. This raises the question of how the argument structures of the two verbs can conflate so as to form the argument structure of the whole 'macro-verb'. A systematic study (François 2004a) has shown Mwotlap to follow strict rules in this regard: thus, (13) below will show how the combination of  $V_1$  'x punch' and  $V_2$  'y cry' regularly results in a transitive macro-verb  $V_1 + V_2$  'x punch-cry y', with the subject of  $V_2$  becoming the object of the serial verb  $V_1 + V_2$ .

The results of this study can be stated, following a *subject-V-object* convention, with simple formulas of the type:  $[x - V_1 - y + y - V_2 = x - [VP] - y]$ . These argument-fusion rules can in turn be grouped in a simple chart (Table 2). The two rows state whether  $V_1$  is intransitive ( $x - V_1$ ) or transitive ( $x - V_1 - y$ ); the eight columns not only show the transitivity value of  $V_2$ , but also the identity of arguments involved ( $x, y, z$ ).<sup>3</sup>

Due to lack of space, we will not illustrate each of these combinations in detail,<sup>4</sup> and will only present the major observations with regard to argument sharing rules.

### 3.2. SUBJECT SHARING PRINCIPLES

As one would expect, it is common for two serialized verbs to share their subject:

- (12) Tita [ta-hag dēyē] nēk l-ēm  
Mum FUT-sit expect 2sg in-house  
'Mum will stay at home waiting for you'  
 $\rightarrow \{x-V_1 + x-V_2-y = x-[VP]-y\}$

However, subject sharing is not obligatory in Mwotlap. A clear example of this is the case of 'switch-function serial verbs', in which  $V_2$ 's subject coincides with  $V_1$ 's object:

TABLE 2. The eleven argument-fusion rules for Mwotlap SVCs

	$E-V_2$	$x-V_2$	$y-V_2$	$x-V_2-y$	$z-V_2$	$x-V_2-z$	$z-V_2-y$	$y-V_2-z$
$x-V_1$	$x-[VP]$	$x-[VP]$	$x-[VP]-y$	$x-[VP]-y$				*
$x-V_1-y$	$x-[VP]-y$	$x-[VP]-y$	$x-[VP]-y$	$x-[VP]-y$	$x-[VP]-z$	$x-[VP]-z$	$x-[VP]-z$	*

<sup>3</sup> In these formulas,  $x$  designates the subject of  $V_1$ ;  $y$  is any second argument distinct from  $x$  (either  $V_1$ 's object,  $V_2$ 's subject, or  $V_2$ 's object); finally,  $z$  is any third argument distinct from  $x$  and  $y$  (either  $V_2$ 's subject or object). The mention of  $z$  in a formula is only relevant if it contrasts with both  $x$  and  $y$ , and therefore, if it follows a transitive  $V_1$ ; hence the hatched areas in the chart. The star means 'unattested'. As for the first column of the table, see §4.3.

<sup>4</sup> For those minor patterns which are not exemplified here, see François (2004a).

- (13) Tali [mi-tit teñen] Kevin  
T. PER-punch cry:REDUP K.  
'Tali made Kevin cry by punching him'  
 $\rightarrow \{x - V_1 - y + y - V_2 = x - [VP] - y\}$

But Mwotlap shows certain configurations that are typologically even more original than (13). One of these is a variant of what is known as 'cumulative subject' (see example (34) in the introductory chapter): if one subject semantically includes the other (in the same way as *we* includes *I*), then the subject of the whole SVC will correspond to the more inclusive of these two subjects. This case is best exemplified by sentences meaning 'accompany (someone)', where '( $x+y$ )-go +  $x$ -take- $y$ ' becomes '( $x+y$ )-[accompany]- $y$ ':

- (14) dō [so tatal tēy] no le-tno plēn?  
inc:du PROSP walk hold 1sg LOC-place plane  
'Will you accompany me to the airport?'  
(lit. Shall *the two of us* walk-and-take me to the airport?)  
 $\rightarrow \{(x+y)-V_1 + x-V_2-y = (x+y)-[VP]-y\}$

Interestingly, this original construction has made its way into the Bislama pidgin spoken on Motalava: the equivalent of (14) would be *YUMITU karem MI i go long eapot?* The people of the neighbouring islands, whose Bislama would be slightly different here (*YU karem MI i go...*), are sometimes amused by this strange dual subject, directly calqued from Mwotlap.

The second configuration we would like to mention here is perhaps even more significant, because it contradicts the claim often made (e.g. Durie 1997: 291) that the elements of an SVC must share at least one argument. In Mwotlap, it is not unusual to serialize two verbs having no participant in common at all. This happens typically when  $V_1$  refers to a single-participant action, and  $V_2$  refers to its effect upon another participant. The output of this combination is a transitive macro-verb—as is made clear by (15) and the corresponding formula:

- (15) ige susu [ma-gayka matyak] no  
pl small:REDUP PER-shout be.awake 1sg  
'I was woken up by the kids shouting'  
 $\rightarrow \{x-V_1 + y-V_2 = x-[VP]-y\}$

Note that in this type of sentence, both verbs are intransitive, as they individually refer to single-participant events:  $\{x - V_1\}$  the kids were shouting in the backyard,  $\{y - V_2\}$  I awoke. Yet the serialization of these two intransitive verbs eventually forms a transitive macro-verb  $\{x - [VP] - y\}$ , as though the action now described were that of an agent ('the kids') upon a patient ('me').

Among other attested combinations, we can mention the following. Note that these examples, whatever the ambiguities of translation, all combine intransitive verbs.

- (16) ne-leñ [mi-yip hal-yak] na-kat  
ART-wind PER-blow fly-away ART-cards  
'The wind blew the cards away'
- (17) nēk [so en mōkhēg] nē-kle!  
2sg PROSP lie.down take.rest ART-back:2sg  
'You should lie down (so that) your back can rest!'
- (18) nēk [mi-tig mēlēmlēg] na-lo den kemem  
2sg PER-stand black ART-sun from 1exc:pl  
'Standing as you are, you're hiding the sun from us'  
(lit. You're standing dark the sun from us)

Incidentally, all the examples (15)–(18) form a subcase of what will later be defined as ‘causative serialization’ (§4.2; see Table 4). Its difference with mainstream causatives is the intransitive nature of  $V_1$ , semantically referring to a single-participant event.

### 3.3. OBJECT SHARING PRINCIPLES

Similar remarks can be made on the issue of object sharing. Of course, serialized verbs can share their object, as we saw in (9); but it can also happen that each verb possesses its own underlying object. Because Mwotlap does not allow for ditransitive constructions, only one of these two objects can be retained for the whole SVC, and this may result in syntactic conflicts.

The principle is that the last argument introduced by  $V_2$  ( $z$ ) overrules the object of  $V_1$  ( $y$ ); the latter disappears from the argument structure of the SVC, and can only be retrieved from the context. In (19), the object of  $V_1$  *tow* (*n-eh* ‘song’) is only mentioned in the topic clause:

- (19) nēk so tow n-eh en,  
2sg PROSP compose ART-song ANAPH  
nēk [tow tataq] na-myōs nōnōm  
2sg AOR:compose follow ART-desire your  
'(if) you compose a song, you just compose following your fancy'

There is no place for the argument ‘song’ in the resulting serial structure; the only object retained is the patient of  $V_2$  (‘follow your fancy’):

you COMPOSE song + you FOLLOW fancy = you COMPOSE—FOLLOW fancy  
that is: { $x-V_1-y+x-V_2-z=x-[VP]-z$ }

An even rarer example of this sort of alchemy is provided by the next sentence:

- (20) ige ūmey a kēy so lep n-eh en,  
(pl) REL SUBORD 3pl PROSP take ART-song ANAPH  
nok [se lep] kēy  
1sg AOR:sing take 3pl  
'Those who want to learn the song, I get them to *learn* it by *singing* it'

The pattern here is as follows:

I SING *song* + they LEARN *song* = I SING—LEARN them  
that is: { $x-V_1-y+z-V_2-y=x-[VP]-z$ }

What is perhaps most striking in all these cases of argument restructuring is their perfect regularity (see Table 2). And, in fact, the tighter the syntactic constraints are, the more efficiently they allow the speaker to forge new combinations, and the hearer to interpret them.

## 4. A functional classification of Mwotlap SVCs

The previous sections were essentially dedicated to the formal and structural properties of Mwotlap serial verbs; we will now undertake a brief semantic typology of these constructions. Interestingly, this functional classification will turn out to be strongly linked to the formal one, as the three major functional categories of serial verbs we recognize depend on whether the subject of  $V_2$  is the same as  $V_1$  (‘concurrent’ serialization), whether it is another participant (‘causative’ serialization), or a whole proposition (‘event-argument’ serialization).

### 4.1. CONCURRENT SERIALIZATION

Despite the empirical diversity of serial constructions in Mwotlap, it is possible to identify a first major functional type: this is when  $V_1$  and  $V_2$  refer to two simultaneous facets of a single event, performed by the same subject. This semantic value, which we identify as ‘concurrent serialization’, encompasses a variety of argument structures, with the only proviso that the subject must be the same for  $V_1$  and for  $V_2$  ( $x$ ). The relevant patterns are listed in Table 3.

An illustration of this functional type would be the following:

- (21) kōyō [ma-tatal kaka] le-mtehal  
3du PER-walk chat LOC-road  
'They were discussing while walking along the road'

See also (6) *hohole galgal* /talk lie/; (9) *yoñteg vēglal* /hear know/; (12) *hag dēyē* /sit expect/; (19) *tow tataq* /compose follow/. Whatever their formal and semantic diversity, all these sentences share one essential property: they show the same participant performing two actions ( $V_1$  and  $V_2$ ) at once. Quite originally, the

TABLE 3. The five formal subtypes of ‘concurrent’ serialization

	$E-V_2$	$x-V_2$	$y-V_2$	$x-V_2-y$	$z-V_2$	$x-V_2-z$	$z-V_2-y$	$y-V_2-z$
$x-V_1$	—	x-[VP]	—	x-[VP]-y				*
$x-V_1-y$	—	x-[VP]-y	—	x-[VP]-y	—	x-[VP]-z	—	*

same-subject serial verbs of Mwotlap always refer to concurrent, complementary facets of a single event, never to successive actions.<sup>5</sup>

Of course, nothing prevents the ‘concurrent’ type from being divided into some secondary functional subtypes, depending on the lexical nature of  $V_1$  and/or of  $V_2$ . For example, the combination of any motion verb ( $V_1$ ) with the adjunct *tēy* ‘hold’ ( $V_2$ ) translates the notion of ‘carry, bring’:

- (22) Bōybōy [mē-hēw tēy] me na-mtig  
B. PER-descend hold hither ART-coconut  
'Boyboy has brought the coconuts down'

More generally,  $V_1 + tēy$  receives a comitative reading, as in (1) *lak tēy* /dance hold/ ‘dance with (someone)’, or in (8) *suwyeg-qen tēy* /cast-net hold/ ‘go net-fishing with (my shoes on)’. Another kind of comitative—still a case of ‘concurrent’ serialization—can be formed with  $V_2$ , *b(i)yin* ‘help, join’:

- (23) dō [so lak biyiñ] kēy?  
iinc:du PROSP dance join 3pl  
'Shall we dance with them?'

This broad category of ‘concurrent serial verbs’— $V_1$  and  $V_2$  being two facets of a single predication, with the same subject—also covers more abstract configurations, such as the comparative. This consists of the combination of a stative  $V_1$  plus a verb we have already discussed (§§2.2 and 2.4), *v(e)teg* ‘put down, leave<sub>TR</sub>, get away from, surpass’:

- (24) kē [nē-mnay veteq] nēk  
3sg STAT-clever leave 2sg  
'He's cleverer than you' (lit. he's clever he leaves you behind)

It may seem surprising to group in a single functional category such diverse semantic values as simultaneous action (21), comitative (1, 23), and comparative (24). However, it must be clear that our present objective is not to classify Mwotlap serial verbs according to their translation equivalents in English. Rather, it is to illustrate how a single linguistic device—namely, the mere sequence of two verbs  $x - V_1 \dots + x - V_2$ —can be powerful enough to consistently encode a wide range of semantic relations, which in other languages would have been formally broken down into many distinct structures (adverbs, prepositions, gerunds or converbs, subordinate clauses, etc.).

<sup>5</sup> The only apparent exception to this statement would be the frequent combination of the verb *van* ‘go’ with a second verb (e.g. *van wēl nu-suk* ‘go buy some sugar’), which indeed has a sequential interpretation. In fact, this sequence of two verbs is not a standard case of SVC: first, because the verb following *van* belongs to the inventory of heads ( $V_1$ ) rather than of adjuncts ( $V_2$ ); second, this structure allows the sequence  $V-V$  to be separated by a directional, which is strictly forbidden to genuine serial structures.

TABLE 4. The five formal subtypes of ‘causative’ serialization

	<i>E</i> - $V_2$	$x$ - $V_2$	$y$ - $V_2$	$x$ - $V_2-y$	$z$ - $V_2$	$x$ - $V_2-z$	$z$ - $V_2-y$	$y$ - $V_2-z$
$x$ - $V_1$	—	( $x$ -[VP]- $x$ )	$x$ -[VP]- $y$	—				*
$x$ - $V_1-y$	—	—	$x$ -[VP]- $y$	—	$x$ -[VP]- $z$	—	$x$ -[VP]- $z$	*

#### 4.2. CAUSATIVE SERIALIZATION

A radically distinct configuration is when  $V_2$  refers to the effect of  $V_1$  upon a second participant. This serializing pattern is the only way to form causatives in Mwotlap. Here we are not dealing with simultaneous actions any more, but with a cause–effect relationship, which necessarily implies that  $V_1$  comes before  $V_2$  in time. Once again, this large functional category may encompass several formal structures (Table 4)—the only criterion being that  $V_2$ ’s subject ( $y$  or  $z$ ) be a participant distinct from  $V_1$ ’s subject ( $x$ ).

The most widespread illustration of causative serialization is the so-called ‘switch-function’ SVC, that is,  $\{x\text{-}V_1-y + y\text{-}V_2 = x\text{-}[VP]\text{-}y\}$ . It can make use of two dynamic verbs, as we saw in (13) *tit tenteñ* /punch cry/; but most of the time, the second verb  $V_2$  is a stative verb or an adjective, as in (11) *yim mat* /stone dead/.

The causative function is not restricted to this standard switch-function pattern. In §3.2, we saw how a cause–effect relationship could be expressed by a sequence of two intransitive verbs, the subject of  $V_2$  being absent from the underlying structure of  $V_1$ ; examples such as (16) *yip halyak* /blow fly.away/ may be described as ‘low agency causative serialization’. Finally, a sentence such as (20) *se lep* /sing learn/, despite its structural originality, clearly belongs to the same category of causative serialization.

#### 4.3. EVENT-ARGUMENT SERIALIZATION

The last major functional type that can be identified recalls the role played by English manner adverbs: a stative, intransitive verb  $V_2$  constitutes a comment on the first verb  $V_1$ . The underlying subject of  $V_2$  is not an individual participant, but the whole event (abbreviated *E*) corresponding to  $V_1$ —more precisely,  $V_1$  and its arguments. This definition corresponds to two formal subtypes (Table 5).

TABLE 5. The two formal subtypes of ‘event-argument’ serialization

	<i>E</i> - $V_2$	$x$ - $V_2$	$y$ - $V_2$	$x$ - $V_2-y$	$z$ - $V_2$	$x$ - $V_2-z$	$z$ - $V_2-y$	$y$ - $V_2-z$
$x$ - $V_1$	$x$ -[VP]	—	—	—				*
$x$ - $V_1-y$	$x$ -[VP]- $y$	—	—	—	—	—	—	*

The first case was illustrated by (1) *lak yoyoñ* /dance quiet/ ‘dance calmly’; (5) *hag qaqa* /sit stupid/ ‘stay idly’; or (10) *leg maymay* /married strong/ ‘be fully married’. The second case involves a transitive V<sub>1</sub>:

- (25) na-bago [mi-ñit maymay] na-ñalñal en  
ART-shark PER-bite strong ART-girl ANAPH  
‘The shark bit viciously (*lit.* bit strong) the girl’  
 $\rightarrow \{x\text{-V}_1\text{-y} + e\text{-V}_2 = x\text{-[VP]-y}\}$

In this sentence, the subject of *mymay* (‘be strong’) is not the shark, let alone its victim; what is meant to be ‘strong’ is event V<sub>1</sub> itself, that is, ‘the shark’s biting the girl’. Formally speaking, one will notice that this sort of serialization always leaves the argument structure of the head intact.

This construction explains why the lexicon of Mwotlap almost lacks manner adverbs (like Eng. *strongly, gently*): this role is played by adjectives in the adjunct position, in the structure we call ‘event-argument serialization’.

## 5. Multiverb serialization

All the rules we have seen, whether formal or functional, have been illustrated by serial verbs with only two members. The last issue we would like to address concerns multiverb serialization. How can the speaker calculate, say, the argument structure of an SVC with four elements? This problem is easily solved if one remembers that any macro-verb resulting from serialization behaves exactly like a simple verb; it then becomes possible to describe any string of verbs starting from the head (V<sub>1</sub>), moving rightwards, and recursively applying the rules defined for any pair of verbs:

$$\{[(V_1-V_2)-V_3]-V_4\}$$

We will illustrate this principle with two examples.

- (26) kēy [lañ mat veteg] hōw nō-lōmgep en  
3pl AOR:beat die leave down ART-boy ANAPH  
‘They got rid of the boy by beating him to death’

The first combination is a CAUSATIVE SVC, *lañ mat* /beat die/ ‘beat to death, kill’:

$$\{they\text{-beat-boy} + boy\text{-dead} = they\text{-(kill)-boy}\}$$

This macro-verb is then incorporated into a CONCURRENT SVC, *lañ-mat veteg* / (kill) leave<sub>TR</sub> / ‘get rid of (someone) by killing him’, which forms a transitive ‘super-macroverb’:

$$\{they\text{-(kill)-boy} + they\text{-leave-boy} = they\text{-(get.rid.of)-boy}\}$$

Finally, we can now fully analyse the complex example in the first page of this chapter:

- (1) [lak tēy yoyoñ ēwē] no  
AOR:dance hold be.quiet be.fine me  
‘Just dance with me calmly’

This string of four verbs must be analysed step by step. The intransitive verb V<sub>1</sub> ‘dance’ and the transitive V<sub>2</sub> ‘hold’ together form a case of CONCURRENT SVC; the result is a transitive verb with a comitative reading:

$$\{you\text{-dance} + you\text{-hold-me} = you\text{-(dance.with)-me}\}$$

In a second stage, this macro-verb is embedded in two successive EVENT-ARGUMENT SVCs, which leave its argument structure intact:

$$\{you\text{-(dance.with)-me} + it\text{-is.quiet} = you\text{-(dance.calmly.with)-me}\}$$

$$\{you\text{-(dance.calmly.with)-me} + it\text{-is.fine} = you\text{-(just.dance.calmly.with)-me}\}$$

The pronoun *no* in (1) is both the object of V<sub>2</sub> ‘hold’ and the object of the whole serial verb construction, which ultimately behaves as a single verb.

The general tendency, as illustrated in this example, is for event-argument SVCs to occur towards the end of the VP—that is, they form the final steps in the chronology of multiverb serialization. The two other types of SVCs are typically met at the beginning of a serial string, with both orders attested equally: either Causative embedded in Concurrent (26), or the reverse.

## 6. Conclusion

Verb serialization is perhaps one of the domains of Mwotlap grammar which are the most productive and subject to historical change. Certain unattested combinations may one day come to light, while other sequences will eventually disappear; some verbs acquire novel properties according to their position as a head or an ‘adjunct’; some lexemes even progressively change their categorial status and specialize in the function of modifier, either grammaticalizing as a valency-increasing applicative, or simply becoming some sort of adverb, a new building block for phraseological innovation. But paradoxically, although this evolution derives intricately from the formal and functional properties of verb serialization, methodological concerns make it difficult to integrate them into the description of ‘serial verbs’ strictly speaking (§1.3).

Language typology tends to focus more on ‘universal’ word classes such as verbs, at the risk of leaving certain language-specific categories undescribed, like the one we called ‘adjuncts’ in Mwotlap. And yet, studying this rich class of verb modifiers, many of which originally come from verbs in former SVC patterns,

would logically constitute the next step in the description of Mwotlap verb serialization.

## References

- BRIL, I. and OZANNE-RIVIERRE, F (eds.) 2004. *Complex predicates in Oceanic languages: Studies in the dynamics of binding and boundness. Empirical Approaches to Language Typology*. Berlin: Mouton de Gruyter.
- CROWLEY, T. 1982. *The Paamese language of Vanuatu*. Canberra: Pacific Linguistics.
- 1987. 'Serial verbs in Paamese', *Studies in Language* 11.35–84.
- DURIE, M. 1997. 'Grammatical structures in verb serialization' pp. 289–354 of *Complex Predicates*, edited by A. Alsina, J. Bresnan, and P. Sells. Stanford: CSLI.
- FRANÇOIS, A. 2000. 'Vowel shifting and cloning in Motlav: historical explanation vs. formal description' pp. 49–68 of *Proceedings of AFLA 7 (The Seventh Meeting of Austronesian Formal Linguistics Association)*, edited by M. Klamer. Amsterdam: Vrije Universiteit.
- 2001. *Contraintes de structures et liberté dans la construction du discours. Une description du mwotlap, langue océanienne du Vanuatu*. Ph.D. thesis, Université Paris-IV-Sorbonne.
- 2002. *Araki: A disappearing language of Vanuatu*. Canberra: Pacific Linguistics.
- 2003. *La sémantique du prédicat en mwotlap (Vanuatu)*. Collection Linguistique de la Société de Linguistique de Paris, 84. Leuven, Paris: Peeters.
- 2004a. 'Chains of freedom: Constraints and creativity in the macro-verb strategies of Mwotlap', in I. Bril and F. Ozanne-Rivierre (2004), pp. 107–43.
- 2004b. 'La réduplication en mwotlap : les paradoxes du fractionnement', pp. 177–94 of *Les langues austronésiennes*, edited by E. Zeitoun. *Faits de langues* n°24. Paris: Ophrys.
- HYSLOP, C. 2001. *The Lolovoli dialect of the North-East Ambae language, Vanuatu*. Pacific Linguistics, 515. Canberra: Australian National University.



## Noun articles in Torres and Banks languages: Conservation and innovation

*Alexandre François*

### **1. Introduction**

In his article “Common noun phrase marking in Proto Oceanic” (1985a), Terry Crowley described the various ways in which the common noun article \*a/\*na evolved across the modern languages of Oceania. While some languages are conservative, others have lost all traces of the original NP marker, or have only retained it partially. Quite often, he showed, what was once a fully productive article became accreted to the noun root, progressively losing its syntactic status as an article.

The languages of Vanuatu are sometimes described as having essentially followed this evolution pattern of article loss, whether involving accretion to the noun root or not. Lynch (2001b) says: “Very few languages of North and Central Vanuatu have articles per se”, and Lynch, Ross and Crowley (2002: 38): “What was historically an article has in many of the languages of Vanuatu (...) been fused with the noun root, being morphologically inseparable in all, or at least most, morpho-syntactic contexts.”

While such statements are certainly true for other languages of Vanuatu, I will show in this paper that they hardly represent the two northernmost island groups of the archipelago, the Torres and Banks Islands (a province usually abbreviated as “Torba”). The seventeen languages spoken in this area (Figure 1), which were still little known until recently,<sup>1</sup> present diverse but essentially similar systems of noun articles. Overall, they show few examples of complete article loss, and only four genuine cases of article accretion. On the contrary, what I observed is that Torba languages still make regular use of noun articles in a way reminiscent of their POc ancestor, albeit in a different manner.

Remarkably, this was Terry Crowley’s own insight when he drew his map of article retention in Oceania (1985a: 162), despite terrible gaps in language documentation in those times. In a way, the present study can be seen as an occasion to confirm and refine his correct intuition, by making available the first-hand data he didn’t have access to twenty years ago. In addition, I will include here information on other articles, and describe certain patterns of innovation, whether

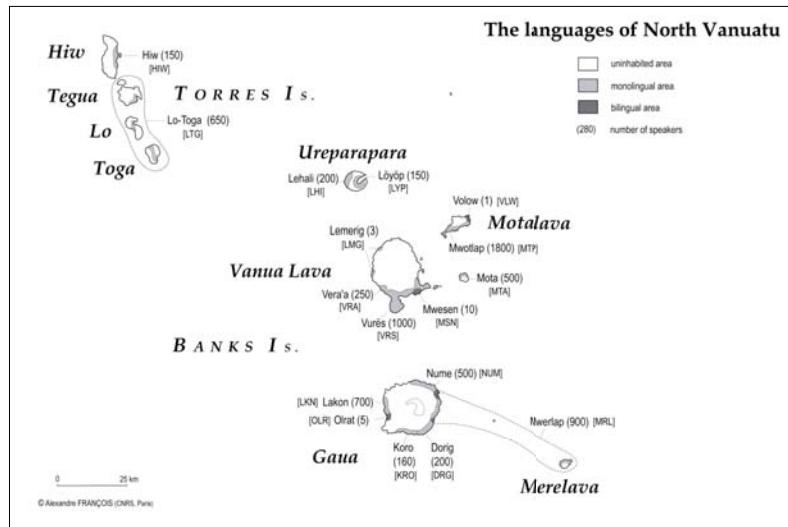


Figure 1. The languages of North Vanuatu

morphological or syntactic, that these Torba languages have gone through.

After describing the form of NP articles (section 2), I will analyse their distribution across noun categories (section 3), and finally discuss their syntactic functions (section 4).

## 2. The morphology of articles

The basic pattern attested in north Vanuatu, as in various other parts of Oceania, is for nouns—whether definite or not—to require a preposed article in order to form a valid NP (see section 4). Thus in Lo-Toga, common nouns must be preceded by the invariable article /na/: e.g., *na rɔ̄ya* ‘a/the tree, Ø/the trees, Ø/some/the wood’.<sup>2</sup>

For reasons of length, I will choose to focus my study on major NP articles, that is, those shared by entire categories of nouns in each language, and made obligatory in most syntactic contexts.<sup>3</sup> After an inventory of these articles (section 2.1), I will describe the diversity of their morphological statuses (section 2.2).

### 2.1 Inventory of articles across Torba languages

The choice of articles is governed by the semantics of the following noun, according to such criteria as *singular vs plural; common vs personal; directly possessed*

Table 1. Inventory of noun articles in the Torba languages

abbr.	language name	directly possessed common N	other common N	human personal singular N	human plural N
HIW	Hiw	nə		Ø	təkŋʷa
LTG	Lo-Toga	nə		Ø	hakʷərə
LHI	Lehali	n-		Ø	kʷɔy
LYP	Löyöp	n-		Ø	de (?)
VLW	Volow	n(V)-		Ø	iŋε
MTP	Mwotlap	nV-		Ø [i/]	iŋε
LMG	Lemerig	n-		i	iŋyε
VRA	Vera'a	(i)n [n/]		ɛ	ɛ raya
VRS	Vurës	na	ɔ	i	i reyε
MSN	Mwesen	ɔ		ɛ	ɛ ra
MTA	Mota	na	o	i	i rat(yai)
NUM	Nume	na-	u [w/]	Ø	ra mel
DRG	Dorig	na	ɔ	i	i ra
KRO	Koro	na	ɔ	i	i rat
OLR	Olrat	Ø [n/]		Ø	nty wu
LKN	Lakon	(i)n		i	yi: wi:
MRL	Mwerlap	nV-		i	re

vs *non-directly possessed* nouns (see section 3). A comparative list of these articles is proposed in Table 1.<sup>4</sup>

As Table 1 shows, the number of articles in each language is variable. Olrat has basically no article at all, except for the plural marking on human nouns. Vurës, Mota, Dorig and Koro possess as many as four distinct articles. Other languages have either two or three.

The various forms listed here may be reduced to a small number of etyma (see Lynch, Ross and Crowley 2002: 71).

The POc article \*a/\*na is reflected—always under its form \*na—in essentially all the languages of the area. Most often it is still a productive full-fledged article, whether a clitic or a prefix. In Olrat, it is only reflected as an accretion on certain nouns. Remarkably, Vera'a reflects it both as an accretion and as a free article (section 2.2.2). Only Mwesen has lost all traces of \*na (section 3.1).

Six languages make use of an article of the form ɔ, o or u. Available evidence suggests it should be reconstructed as \*wo.<sup>5</sup> This article \*wo is obviously an innovation, both in formal and functional terms: see section 3.1.

The POc personal article \*i/\*e is represented in as many as ten languages, reflecting either \*i or \*e. Out of these ten reflexes, nine are still productive, while one (Mwotlap) only exists as an accretion on certain nouns (section 2.2.2). Most languages have lost \*i/\*e altogether: their personal NPs take zero article. Note that no language retains the other POc personal article \*qa.

Finally, the maximum formal diversity is found with human plurals. However, knowledge of the regular phonetic correspondences throughout the Torba

area (François 2005b) suggests all these attested forms—except HIW *təkyl'a* and LYP *de*—include a reflex of a syllable \*ra, albeit in a hidden way (e.g., LKN *yɪr* < \**yɪr* < \**kira*). This obviously corresponds to POc \*ra marking 3rd person (normally human) plural. In six languages (Volow, Mwotlap, Lemerig, Vera'a, Vurës, Mota), the form regularly reflects a phrase \*i/e ra kai, which can be analysed as Personal \*i + plural \*ra + POc \*kai ‘native, person’ (Pawley 1976): e.g., VLW/MTP *iye* < *iyye* < *iryē* < \*i-rayai < \*i ra kai.

## 2.2 Morphological status

### 2.2.1 From clitic to prefix

All the articles cited in Table 1 are immediately followed by the noun which is the head of the NP. Not surprisingly, the two morphemes together form a prosodic unit. The whole noun phrase follows a single intonational contour, whereby the article lacks its own primary stress, being prosodically integrated to the following noun: e.g., LTG *nə=yə'huwa* ‘rat’. This status of noun articles as proclitics, which must probably be reconstructed also for POc (e.g., \*na=ka'supe), is still well attested in Torba languages. For example, all the human articles—except obviously in their accreted versions—whether singular or plural, still behave as clitics; and so do the reflexes of \*wo. But the situation is more complex regarding \*na.

In six languages (Hiw, Lo-Toga, Vurës, Mota, Dorig, Koro), \*na still shows all the properties of a clitic. In Vera'a and Lakon, this clitic has the basic form /in/, but undergoes inverse elision (aphaeresis) after a vowel-ending word: compare VRA *yən m̥ mes* ~ LKN *yən m̥ mæh* ‘eat a fish’ with VRA *lə\_n mes* ~ LKN *lae\_n mæh* ‘take a fish’. Despite this rule of sandhi depending on the preceding word, the syntactic scope of this article /in/ is clearly the following noun: it can still be described as a proclitic.

In seven other languages (Lehali, Löyöp, Volow, Mwotlap, Lemerig, Nume, Mwerlap), the prosodic incorporation of the article eventually triggered also its phonological integration to the following noun, so that it must now be considered a prefix. There are three ways for the article to form a single phonological word with the following noun (see François 2000; 2005b):

1. VOWEL HARMONY [Volow, Mwotlap, Mwerlap]: the vowel of \*na regularly undergoes partial or complete assimilation to the first vowel of the noun: e.g., \*na *kutu* ‘louse’ > VLW/MTP *ni-yit*.
2. VOWEL ELISION [Lehali, Löyöp, Volow, Lemerig]: the vowel of \*na is deleted, so that the article is reduced to the consonant *n*. Whereas other languages only allow this elision before another vowel (e.g., \*na *ikan* ‘fish’ > LTG *n' iŋə* ~ MRL *n-eay*), these four languages also allow it before a consonant (e.g., \*na *patu* ‘stone’ > LMG *n-ve?*).

3. PHONOTACTIC INTEGRATION [Mwotlap, Nume, Mwerlap]: \*na is regularly followed by consonant clusters, which otherwise never occur word-initially in the language. Thus in Mwotlap, compare the form with article *na-ɣhow* ‘rat’ < \*na *kasupe* with the bare noun *ɣhow* showing epenthesis.

If an article meets any of these criteria, then it is a prefix. Conversely, when the phonological form of the article and that of the noun are independent from each other (e.g., LTG *nə=yə'huwa* ‘rat’), one may still speak of a clitic.

Crucially, the change in morphological status—from clitic to prefix—doesn’t necessarily involve any change in syntactic behaviour. Thus, the prefix *nV-* in Mwotlap can still be analysed as a genuine article, just as much as its clitic ancestor \*na. The lack of a clear distinction between the morphological and syntactic levels has led certain scholars to confusion. For example, Crowley (2002c: 591) had this erroneous analysis about Mwotlap: “There are no articles in Mwotlap. The original prenominal article /\*na/ has been reanalysed as part of the citation form of the noun.”

Only syntactic properties should define the status of a morpheme as a productive article, such as:

1. its capacity to affect the whole noun lexicon, or at least entire, definable, categories of this lexicon (e.g., inanimate nouns), including new lexical items when they fall into these categories
2. the existence of productive rules governing the presence vs absence of this article, depending on the syntactic context.

Sections 3 and 4 of this chapter will demonstrate that these two requirements can be fulfilled by a prefix just as much as a clitic.

### 2.2.2 Article accretion

The historical phenomenon of “article accretion” brings about a different situation altogether. In this case, what was once a free article has been attracted not only to the following phonological *word*, but to the following *root* itself, to such an extent that it has lost its status as an article. While this process seems to have occurred quite often throughout Vanuatu (Crowley 1985a; Lynch 2001b) and in New Caledonia (Ozanne-Rivierre 1992), in the Torba area it is only attested in a few cases.

In Vera'a and Olrat, the article \*na has been accreted to vowel-initial noun radicals. To take reflexes of POc \*(na) quraj ‘lobster’, one must carefully distinguish between, on the one hand, such forms as MTP *n-iy* ~ MRL *n-ʊər* where the article is still syntactically a free prefix; and, on the other hand, VRA *n-irɪ* ~ OLR *n-ʊrɪ*, where it has become an inseparable part of the noun radical.<sup>6</sup> Examining such forms in the light of the two criteria stated above (end of section 2.2.1) makes it clear we are no longer dealing with articles. First, this

consonant /n/ does not affect productively any noun category that would be definable in synchrony, but is arbitrarily present in certain items of the lexicon. Second, there is essentially no syntactic context that allows for regular deletion of this /n/. These words now behave exactly the same as any \*n-initial noun. For example, and quite remarkably, Vera'a allows them to take the common article *m*, also a reflex of \*na —e.g., *m n/ir* 'a/the lobster'.

Similar cases of accretion occurred with other articles. In Nume, vowel-initial noun radicals also underwent article accretion, but this time with \*wo, under the form /w/: e.g., NUM *wew* 'fire' < POC \*api; *wow* 'turtle' < PNCV \*?avua; *wry* 'fish' < POC \*ikan; *wem* 'mat bed' < POC \*qebal; *wim* 'earth oven' < POC \*qumun; *wor* 'lobster' < POC \*quranj; *wak* 'canoe' < POC \*waga. In this case too, the accreted /w/ may co-occur with the free article *u*, itself a reflex of \*wo: e.g., *u w/ur* 'a/the lobster'.

Finally, Mwotlap only reflects the personal article \*i as a fossil vowel at the beginning of about twenty human nouns, essentially kinterms and a few proper names (François 2001: 208–213): MTP *ihi-k* 'my brother' < POC \*i taci-gu; *imam* 'Dad' < \*i mama; *Ik̪ʷet* 'Iqet, name of a cultural hero' (MTA *i K̪ʷat*)—see also *iye* 'plural article' (section 2.1).

### 3. Noun categories and their articles

Section 2 has shown both the unity and diversity of articles across the Torba area, at least regarding their form. As for the distribution of articles across noun categories in each language, it also shows some variety, yet allows for a general description. I will first describe the contrast between \*na and \*wo in the languages that have both (section 3.1), and later will delineate the categories of common vs personal nouns (section 3.2).

#### 3.1 A specific article for inalienable possession

Table 1 showed the existence of an article \*wo in six geographically adjacent languages. Not only is this form unknown outside this small “central Banks” area; but its functional distribution is also, to my knowledge, unusual both from an Oceanic and from a typological perspective. It clearly results from a local innovation, which either never took place in the neighbouring languages, or did and was later reversed.

In five of these languages (Vurës, Mota, Nume, Dorig, Koro), the article \*na has been restricted to only one category of nouns, namely, those [-human] nouns that are inalienably possessed, i.e., directly followed by a possessor (whether a suffix or an NP). The innovative article \*wo is used otherwise, that is, with [-human] nouns that are *not* directly possessed: contrast VRS *na yærvii-k* 'my house' vs *ɔ yøvür* 'a/the house'.

To be precise, a noun normally combines with *na* only if it has a specific human possessor. Thus compare for Dorig:

- Dorig
- |        |                          |            |                   |                            |                            |
|--------|--------------------------|------------|-------------------|----------------------------|----------------------------|
| (1) a. | <i>na</i>                | <i>ssa</i> | <i>i</i>          | <i>vvi-k</i>               |                            |
|        | ART.POSS                 | name       | ART:PERS          | mother-1SG                 |                            |
|        | 'my mother's name'       |            |                   | (possessor [+spec] [+hum]) |                            |
| b.     | <i>ɔ</i>                 | <i>ssa</i> | <i>rkpʷa</i>      |                            |                            |
|        | ART                      | name       | woman             | (possessor [-spec] [+hum]) |                            |
| c.     | <i>ɔ</i>                 | <i>ssa</i> | <i>wasjin nej</i> |                            |                            |
|        | ART                      | name       | place             | that                       | (possessor [+spec] [-hum]) |
|        | 'the name of that place' |            |                   |                            |                            |

Likewise, when an inalienable noun is marked for a generic possessor, it must take the \*wo article: e.g., MTA *na pane-ȳmʷa* 'your hand' vs *o pane-i* 'one's/a hand'.

Furthermore, each of these languages has between four and six possessive classifiers, which behave as a subclass of inalienable nouns. Since their function consists in indexing a possessor, they almost always occur with the article *na*: MTA *na ma-k* 'my X (Drink possession)'; *na ȳmʷ-o-ȳmʷa* 'your X (General possession)'. Remarkably, the noun X which is modified by this classifier, although it is semantically possessed, must bear the \*wo article, because it does not receive itself the possessive morphology: it is *indirectly* possessed. In these five languages, this regularly results in quite paradoxical NP structures, where the noun combines with \*wo while its classifier takes \*na:

- Mota
- |     |                |               |           |                      |  |
|-----|----------------|---------------|-----------|----------------------|--|
| (2) | <i>o</i>       | <i>tkpʷei</i> | <i>na</i> | <i>ȳmʷo-ra</i>       |  |
|     | ART            | garden        | ART.POSS  | POSS.CLF.general-3PL |  |
|     | 'their garden' |               |           |                      |  |

- Vurës
- |     |              |                      |          |           |  |
|-----|--------------|----------------------|----------|-----------|--|
| (3) | <i>na</i>    | <i>ka-ŋ</i>          | <i>ɔ</i> | <i>ak</i> |  |
|     | ART.POSS     | POSS.CLF.vehicle-2SG | ART      | canoe     |  |
|     | 'your canoe' |                      |          |           |  |

Obviously, the languages that only retain \*na have no such asymmetry. Thus the equivalent of (2) in Volow would be *n-tigbʷi nɔ-yɔ-y*, with two instances of \*na.

In summary, these five languages allow \*na only with a suffixable word—whether a noun or a classifier—that is directly possessed, either followed by a personal suffix or a semantically [+human] [+specific] NP possessor. In all other cases, the article is \*wo.

Finally, out of the six languages that reflect \*wo, Mwesen is original in having lost all traces of \*na, and generalised \*wo to all contexts. Thus Mwesen has *ɔ iy̪mʷ* ‘a house’; *ɔ iy̪mʷɔ-k* ‘my house’; *ɔ sa-n ε ritnɔ-k* ‘my mother’s name’; *ɔ pini-m* ‘your hand’; *ɔ mɔyɔ-nir ɔ tukpʷi* ‘their garden’.

### 3.2 Personal vs common articles

Another functional notion that proves relevant for the description of NP articles in Torba languages, as indeed elsewhere in Oceania (Pawley 1972: 32), is the contrast between *personal* and *common* NPs. Formally speaking (Table 1), personal NPs are either marked by \*i/\*e or by zero; they contrast with common NPs, which take \*na or \*wo.

On the semantic level, a “personal” NP normally has a *human specific* referent. This is typically the case with proper names, kinterms, pronouns or deictics with human reference. Thus one finds MSN *ε Tevit* ‘David’; KRO *i mam* ‘Dad’; VRS *i kpʷælyæ-k* ‘my father-in-law’; MTA *i nau* ‘I (1sg pronoun)’; VRA *ε si* ‘who?’; DRG *i at nep* ‘the one there’. The plural articles given in Table 1 often include the personal article \*i/\*e, because they are only used with human referents: MSN *ε tese-n* ‘his brother’ → *ε ra testese-n* ‘his brothers’. Non-human nouns never take a personal article, except in stories where animals or objects are personified: MSN *ε yosow min ε yoto* ‘Rat and Hermit-Crab’.

To be precise, the contrast personal vs common somehow constitutes a semantic continuum, of which only the two ends are clearly defined. On the one hand, proper names, or kinterms with individual reference, must be treated as personal; on the other hand, non-human NPs must be treated as common. The situation is less clearcut in the intermediate zone—that is, non-kin human noun phrases. As a tendency, a given noun will be treated as personal if it points to a definite individual in the given context—e.g., MRL *i vatɔy* ‘the teacher’, DRG *i mayte* ‘the (aforementioned) old lady’, MSN *ε maranay* ‘the chief’, VRS *i biryi-k* ‘my partner’—but as common if it points to a generic or indefinite referent, or to the notional quality of the noun (e.g., predicate ‘be a N’):

Dorig

- (4) *na m-tek ɔ mayte s-ru.*  
1SG PRF-see ART old.woman NUM-two  
'I saw two old ladies.'

Mwesen

- (5) *ε si ɔ maranay εlile?*  
ART:PERS who ART chief here  
- *ε maranay ε no.*  
ART:PERS chief ART:PERS 1SG  
'Who's (the) chief here? – The chief, that's me.'

No clearcut principle can really be asserted here. First, discrepancies are common, whereby the same noun can equally be treated as personal or common—including in the same sentence (e.g., MRL *i bulsala-n* ~ *nu-bulsala-n* ‘her boyfriend’).

Second, certain human nouns appear to be just incompatible with the personal article, whatever their actual reference. This is especially the case of the four common nouns ‘person’, ‘man’, ‘woman’, ‘[non-relational] child’, perhaps because these lexical items are statistically most often used with non-referential or qualitative value—whether as a generic NP, a predicate or an attribute. These nouns are systematically treated as common in all Torba languages, including when they clearly designate a specific individual:

Dorig

- (6) *i ntu-ŋ neŋ sa, ɔ iy̪mʷerat sa...*  
ART:PERS son-2SG that there ART man there  
'that son of you there, the boy there...'

In other languages, this structural asymmetry takes the form of a contrast Ø (personal) vs \*na (common):

Volow

- (7) *(Ø) niti na, gbʷε n-tajmʷan na...*  
ART:PERS son:2SG there that.is ART-male there  
'that son of you there, the boy there...'

Although these four exceptional nouns always take a common article in the singular, they become compatible again with personal marking in the plural: e.g., VRS *ɔ iy̪mʷirj̪mʷiar* ‘a child’ → *i reye iy̪mʷirj̪mʷiar* ‘children’. Finally, a similar paradox can be illustrated with the phrases meaning ‘my wife’. When a language possesses a dedicated kinterm, then it is encoded as Personal: VRS *i yünə-k*; MTA *i rasoə-k*; MRL *i rənatə-k*. Otherwise, a periphrasis will be used with the noun ‘woman’—in which case common articles are required: KRO *na mu-k ɔ rakpʷa*; HIW *na yəkʷen əkiə* (lit. ‘my woman’).

In sum, personal articles (\*i/\*e or zero) are restricted to highly individuated human referents. Common articles (\*na or \*wo) are required in all other cases: that is, for non-human, non-specific, poorly individuated referents; and by extension, with certain nouns that are statistically seldom referential—even when they actually are.

### 3.3 Synthesis

This section has delimited the various noun categories that are relevant to explain the distribution of major articles (\*na, \*wo, \*i/\*e, zero) in Torba

languages. One particular semantic class deserves attention here, namely that of *human specific, highly individuated referents*, as opposed to all the rest (François 2005a). Not only does this feature account for the contrast between personal and common articles; but it also helps define the type of possessor that requires \*na vs \*wo in those languages that possess two common-noun articles. The architecture of the article system in Torba languages—or better, of those systems with the maximum number of distinctions—is summarised in Figure 2.

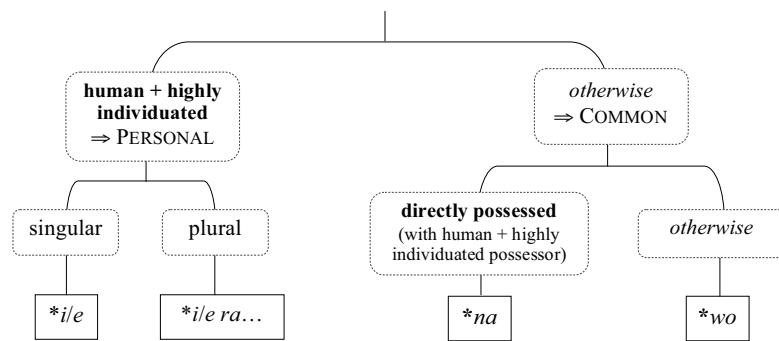


Figure 2. The maximal architecture of Torba article systems

#### 4. The syntax of noun articles

The preceding sections have examined the form and nature of noun articles attested in the Torba area, as well as the semantic properties of nouns that are canonically associated with them. Each NP was cited with its article, as though it just had to be there. In reality, the status of these morphemes as true articles entails the possibility of observing which contexts require their presence vs their absence (see section 2.2.1). This issue should help work out the syntactic *raison d'être* of these articles.

##### 4.1 Syntax of the personal article

Even if a noun fulfills all the semantic requirements to be compatible with a personal (rather than a common) article, this doesn't mean this article will always be there.

First, most languages seem to treat the \*i/\*e article as more or less optional, even in those contexts where its presence would be expected. For example, in

Dorig, ‘their grandfather’ in subject position is normally *i tbirar*, but my oral corpus shows several instances of just *tbirar* without the article.

Furthermore, while most syntactic functions require—or at least are compatible with—the personal article (subject, direct object, object of a preposition, possessor, predicate), at least one excludes it: the vocative. Following the framework developed by Lemaréchal (1989: 37) based on Tagalog data, one could analyse these patterns in terms of change in syntactic status (François 2001: 209):

1. Considered in its bare form, the status “personal phrase” is compatible with just one function—**address**: KRO *tsi-k!* ‘brother!’, MSN *mam!* ‘Dad!’, MTA *Teri!* ‘Terry!’.<sup>7</sup>
2. The personal article \*i/\*e then has the power to “transfer” (Fr. “translate”) this phrase towards a new status, that of a **referential** phrase: KRO *i tsi-k* ‘my brother’, MSN *ɛ mam* ‘Dad’, MTA *i Teri* ‘Terry’. This status allows it to play all the syntactic functions that are open to referential phrases (subject, object, possessor, etc.).

##### 4.2 Syntax of the common articles

Establishing a similar rule for the common articles (\*na/\*wo) is slightly more difficult, due to the diversity observed from one language to another. Especially, languages appear to vary again as to how optional the presence of the article is in those contexts where it is allowed. For example, LYP *n-*, NUM *u*, VRA/LKN (*i*)*n*, MRL *nV-* are dropped at will in spontaneous speech, even when they could be present. In contrast, the rules regarding HIW/LTG *nø*, MTP *nV-*, MSN/VRS/DRG *ɔ...* are much more constraining: basically, if the article is allowed by the context, then it must be there. Obviously, the latter languages are a more reliable source of observation regarding rules for the presence/absence of the common article.

###### 4.2.1 When is the article excluded?

For the sake of consistency, I will cite here data from one language, Mwotlap (François 2001: 187–214), taking it as essentially representative of the whole area. Incidentally, this choice constitutes a belated response to Crowley’s regret (1985a: 161), that “there is no evidence concerning the separability of this prefix [*nV-*] in Motlav [Mwotlap]”.

I will choose the word *nr-bi* ‘fresh water, river’ as illustrative. Although corpus-based statistics (François 2001: 204) show that 73% of this word’s occurrences—as well as its citation form—include *nV-*, many syntactic contexts actually require the noun in its bare form.

X modifies another noun

- (8) a. X indicates the contents or the substance of N<sub>1</sub>:  
*na-plastik bɪ* ‘bottle of water’; *na-kpʷlis bɪ* ‘water puddle’
- b. X is a distinctive property of N<sub>1</sub>:  
*na-mya bɪ* ‘river eel’; *na-pno bɪ* ‘river island’
- c. X is the “possessor” of an alienable noun N<sub>1</sub>:  
*na-mne bɪ* ‘the taste of the water’; *nu-toti bɪ* ‘the river’s source’
- d. X is a “possessed” noun following its classifier:  
*ne-me-k bɪ* ‘my water (to drink)’; *minɔ bɪ* ‘my water (for non-drink uses)’
- e. X complements a dependent noun N<sub>1</sub>:  
*na-mathey bɪ* ‘thirst’ (literally, craving for water)

X modifies a dependent morpheme

- (9) a. X complements a prenominal element:  
*babahne bɪ* ‘the last river’; *na-han bɪ* ‘which river?’;  
*ni-tiy bɪ* ‘genuine water’
- b. X complements a preposition (free or prefixed):  
*ləs bɪ* ‘inside the water’; *taval bɪ* ‘across the river’;  
*(sisyɔy) l- bɪ* ‘(fall) in the water’; *(vnyyil) bɪ- bɪ* ‘(argue) about water’
- c. X complements the linker /ne/ ‘of’:  
*na-mtehal ne bɪ* ‘the course of the river’
- d. X complements the partitive /te/ ‘some’:  
*nɔk sɔ in te bɪ* ‘I want to drink some water’

X modifies a verb or a predicate

- (10) a. X is a non-referential incorporated object within a verb:  
*inin bɪ* ‘drink water’; *haha bɪ* ‘draw water (from a well)’
- b. X is a non-referential incorporated object in a noun compound:  
*tytyty- bɪ* ‘healer’ [lit. ‘water-holder’];  
*ne-men inin- bɪ* ‘dragonfly’ [lit. ‘water-drinking insect’]
- c. X is the (non-patient) internal complement of a verb:  
*vihyi bɪ* ‘be changed into water’
- d. X is the internal complement of an existential predicate:  
*tateh bɪ* ‘there’s no water’; *takpʷse bɪ* ‘there are many rivers’

Others

- (11) X is a TAM-marked predicate noun:  
*n-as mal bɪ lɔk* ‘the ice has [become] water again’

#### 4.2.2 Function of the common noun article

The syntactic contexts listed above share certain essential properties. Typically, the noun appears unprefixed when it constitutes a phrase-internal modifier,

pointing semantically towards a generic notion (‘water’) or a quality (‘watery’) rather than designating a referential entity.

In contrast, the article is required whenever the noun is syntactically the head of an autonomous constituent, whether a verb’s argument, a noun predicate, etc. Semantically, the function of this article consists in embodying the noun’s quality into a discrete, specific referent: *ni-bɪ* ‘some/the water: a specific quantity of water, a river...’

Following the analysis I proposed for the personal article (section 4.1), common-noun articles may be described as a device used to transform **qualifying** noun phrases (MTP *bɪ*) into **referential** noun phrases (MTP *ni-bɪ*). Interestingly, this means that personal and common articles operate upon different input entities (respectively vocative NP vs qualifying NP), but make them converge into the same output (referential NP).

## 5. Conclusion

Overall, Torba languages essentially agree in having kept noun articles alive and productive, as a syntactic device to form referential noun phrases. In each language, selecting the proper article ultimately depends on the syntactic, semantic and pragmatic properties of the NP in its specific context.

The various structures described in this paper may be of some interest to two communities of linguists. Oceanists in quest of historical evidence will notice the retention of several inherited features that have been lost elsewhere in Vanuatu. On the other hand, these formal structures delineate certain cognitive categories that may be of interest to typological linguists, whether they sound universally familiar (e.g., human specific referents) or more unusual (inalienable nouns possessed by a human specific referent). Building bridges between Oceanists and typologists was precisely one of Terry Crowley’s major achievements, and a perspective we shall hopefully continue to bear in mind.

## Notes

1. The data cited in the present paper were collected by the author during four field surveys: May–July 1998 for Mwotlap, Vurës and Mwesen; July–September 2003 for Volow, Vera'a, Nume, Dorig, Koro, Olrat, Lakon and Mwerlap; July–August 2004 for Mota, Lehalì and Lo-Toga; January 2006 for Lemerig, Löyöp and Hiw. Note that I use here the term “languages” in the broad sense of “speech varieties”, regardless of whether some can be grouped together as dialects of a single language.
2. Throughout this article, forms are transcribed phonemically rather than using standard orthographies, to enable comparison. Note that /v/ = [β], and that all voiced stops are prenasalised: /b/ = [mb], etc.
3. I will therefore leave unmentioned here certain (quasi) articles with a more limited scope. For example, (1) certain proper names and kinterms reflect a vestigial feminine “article” \*ra/ro;

(2) Mwotlap has a partitive *te* < \*tewa ‘one’; (3) four Gaua languages possess an indefinite article, grammaticalised from \*tuara ‘other’; (4) Torres languages have a set of human markers, used both as pronouns and as articles; etc.

4. Hyphenated forms indicate prefixes, otherwise the article is a clitic. Forms in square brackets and followed by ‘/’ correspond to reflexes that are only vestigial, taking the form of a phoneme that is now incorporated into certain noun roots (see section 2.2.2).

5. The reasons for reconstructing an initial consonant \*w include: (a) the accretion of /w-/ to certain roots in Nume (section 2.2.2); (b) the existence in Mwotlap of an article-like prefix forming honorific nouns or nicknames, with the form *wo-* (François 2001: 242); (c) the form of the noun article *we* ~ *wu* in the Banks “song dialect” (Codrington 1885: 309), an archaic poetic language common to the whole area.

6. Other examples include: VRA *nev* ‘fire’ < POc \*api; VRA *ner* ‘Casuarina’ < POc \*aRu; VRA *nuwu* ~ OLR *now* ‘turtle’ < PNCV \*?avua; VRA *nijmwi* ‘house’ < POc \*Rumaq; VRA *non* ‘sand’ < POc \*qone; VRA *nur* ‘Spondias cytherea’ < POc \*quRis; OLR *num* ‘earth oven’ < POc \*qumun; VRA *naka* ~ OLR *nak* ‘canoe’ < POc \*waga.

7. These bare noun radicals precisely confirm that KRO *i* and MSN *e* are still synchronically productive articles. In contrast, the fossilised \*i of Mwotlap (section 2.2.2) cannot disappear: e.g., *ithi-k!* ‘brother!'; *imam!* ‘Dad!'; *Ikp<sup>w</sup>et* ‘Iqet'.

## Semantic maps and the typology of colexification

Intertwining polysemous networks  
across languages

Alexandre François  
Lacito (CNRS), Fédération TUL

Building upon the model of Semantic Maps (Haspelmath 2003), which typologists have designed mainly for grammatical semantics, this chapter discusses methodological issues for a model in lexical typology.

By breaking up polysemous lexemes of various languages into their semantic “atoms” or senses, one defines an etic grid against which cross-linguistic comparison can be undertaken. Languages differ as to which senses they *colexify*, i.e., lexify identically. But while each polysemous lexeme as a whole is language-specific, individual pairings of colexified senses can be compared across languages. Our model, understood as an empirical, atomistic approach to lexical typology, is finally exemplified with the rich polysemyies associated with the notion “BREATHE”. Intertwined together, they compose a single, universal network of potential semantic extensions.

**Keywords:** *breathe; colexification; etic grid; lexical typology; methodology; polysemy; semantic maps; sense; soul; spirit*

### 1. General issues of lexical typology\*

At first sight, the capacity of the human brain to detect analogies in one’s environment is infinite, and should logically result in lexical polysemy having no limits. And indeed, the more languages we explore, the more examples we find of unique metaphors and unexpected cases of semantic shift – probably one of the most thrilling mysteries and charms of language discovery. But what generally happens is that we focus our

---

\*I would like to thank Martine Vanhove, Maria Kotjevskaja-Tamm, Sergueï Sakhno and Françoise Rose for their precious comments on a previous version of this article.

attention on the most exotic cases, and overlook the information that is of most interest for the hunter of semantic universals: namely, that a great deal of lexical polysemies are in fact widespread across the world's languages, and, as such, deserve to be highlighted and analyzed.

This observation meets the agenda of lexical typology. Indeed, one of the mainstays of typological linguistics is precisely to show that cross-linguistic variation, far from being random and infinite, can in fact be reduced to a limited range of possible cases. And it is the purpose of this whole volume to show that the search for universals and typological tendencies, which has already proven fruitful in phonological or grammatical studies, may perfectly apply to the study of the lexicon too, provided the specific methodological issues it raises are properly addressed.

Generally speaking, one central issue raised by linguistic typology is the necessity to ascertain the comparability of languages. Languages can only be contrasted with accuracy provided a standard of comparison is proposed, defining the common ground against which commonalities and differences can be observed across languages. Studies in grammatical typology have already begun to identify some of the relevant criteria for the comparison of grammar systems. They consist in the many functional features that emerge out of the observation of actual categories in natural languages: such notions as number, animacy, deixis, telicity, agentivity ..., form a solid functional basis for the cross-linguistic analysis of specific points of grammar. But in the less explored domain of lexical typology, the comparability of languages seems less easy to delineate.

Several reasons may account for this scientific gap. For one thing, there is still the widespread idea that grammars are tidy and regular, while lexicons would be open-ended, exuberant and idiosyncratic. With such a perception, it is deemed unlikely that the typological project might come up with any satisfying generalizations in the lexical domain as much as it does in the observation of grammars. Also, the accurate description of lexical data often requires taking into account the many functional properties of real-world referents, to say nothing of the pitfalls of culture-specific vocabulary; this seems to make the comparative project a difficult challenge.

The aim of this article is to discuss and illustrate the possibility of comparing the world's lexicons, by resorting to a methodological tool which has already proven its efficiency among grammar typologists: *semantic maps*. For a given notion taken as the map's pivot, I will suggest a method for drawing a universal network of potential semantic extensions, following the observation of polysemies attested across the world's languages. A useful concept for this study is the notion of *colexification*, which will be introduced in 3.2. Finally, the last part of this paper will illustrate the potentials of this method, by analyzing the complex semantic network associated with the notion "breathe".

## 2. Ensuring the comparability of lexicons

### 2.1 Monosemy vs. polysemy

The first issue that has to be addressed when studying the lexicon, is the nature of the objects to be compared. Indeed, the comparative project will be directly affected by theoretical choices regarding the nature of the word, whether it is understood as intrinsically monosemous, or capable of genuine polysemy. This discussion relates to ongoing debates (see Geeraerts 1993; Nerlich et al. 2003; Riemer 2005) which I will only mention briefly here in relation to the present discussion.

When Saussure defined the sign as the arbitrary pairing of a form (the "signifier") and a concept (the "signified"), he insisted that each concept can only be characterized negatively, insofar as it contrasts with other words of the same language: "Concepts (...) are purely differential; they are defined not positively by their contents, but negatively by their relationship to the other elements of the system."<sup>1</sup> This conception of semantics has led to the structuralist view that the meaning of a given word in one language will never match exactly the meaning of its most usual translation in another language: its "semantic outline", as it were, is unique to that particular system, and cannot be found identical anywhere else. In such a framework, the very project of a lexical typology, aiming to compare lexicons across languages, seems not only difficult, but simply out of the question.

Directly inherited from this structuralist standpoint is the MONOSEMIST approach, whereby a polysemous lexical unit will be analyzed as fundamentally organized around a unique general meaning; its different attested senses in context are understood as resulting from the combination of that core meaning with the pragmatics of each specific speech situation. Conversely, the POLYSEMIST approach considers the multiplicity of meanings to form an intrinsic property of each polysemous word at the semantic level, with no necessity, or even legitimacy, to reduce this multiplicity to an artificial unity. Several attempts have been proposed to reconcile these two contrary approaches, for example, around the notions of "prototype" (Rosch 1973) or "radial categories" (Lakoff 1987).

It is not the purpose of this article to solve such long-discussed issues. What is relevant here is to underline that each point of view is an attempt to handle the dialectic between unity and multiplicity, which is inherent to the paradox of polysemy. Now, it appears that cross-linguistic comparison can be carried out with more precision if the

1. "Les concepts (...) sont purement différentiels, définis non pas positivement par leur contenu, mais négativement par leurs rapports avec les autres termes du système" (Saussure 1972 [1916]: 162).

facts of polysemy are stated explicitly from the perspective of a multiplicity of senses. The idea that each polysemy is fundamentally underlined by a single abstract meaning, though intellectually appealing it may be, results in definitions that are difficult to apprehend with precision, and to test against actual data. To quote the words of Haspelmath (2003: 214), “general-meaning analyses are not particularly helpful if one wants to know in what way languages differ from each other.”

Whatever theoretical viewpoint one adopts concerning polysemy, the only representation that really allows cross-linguistic comparison is therefore one that explicitly spells out the multiplicity of senses making up a word’s polysemy. The question whether these senses are to be understood as pragmatically defined contextual uses of a central meaning (monosemist approach), or as autonomous components at the semantic level (polysemist approach), is somewhat a secondary issue. What is essential is to find a method that will allow us to describe each polysemous network in the full detail of its internal components.

## 2.2 Overlapping polysemes

A first illustration can be proposed, with the English word *straight*. Roughly speaking, this adjective may be broken into at least the following senses<sup>2</sup> (see 3.1. for a discussion of the method):

<i>(rectilinear)</i>	<i>(a straight line)</i>	<i>(heterosexual)</i>	<i>(gay or straight)</i>
<i>(frank)</i>	<i>(straight talking)</i>	<i>(undiluted)</i>	<i>(straight whisky)</i>
<i>(honest)</i>	<i>(a straight guy)</i>	<i>(directly)</i>	<i>(straight to the point)</i>
<i>(classical)</i>	<i>(a straight play)</i>	<i>(immediately)</i>	<i>(straight away)</i>

Its closest translation in French, *droit*, shows a slightly different range of senses:

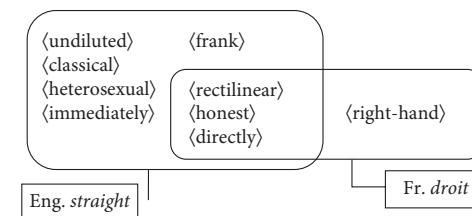
<i>(rectilinear)</i>	<i>(un trait droit)</i>
<i>(directly)</i>	<i>(aller droit au but)</i>
<i>(honest)</i>	<i>(un type droit)</i>
<i>(right-hand)</i>	<i>(le côté droit)</i>

Now, a strictly monosemist approach would probably try to define the core meaning of *straight* by resorting to a general definition, sufficiently abstract so as to encompass all its contextual uses in English. Then it would also propose a unique definition for French *droit*; and because the meanings attested for these two words are so close to each other, it is likely that the two general definitions would end up being quite similar,

2. Throughout this paper, angled brackets (...) are used to represent senses, insofar as they form an element of a polysemous network.

and therefore unable to grasp clearly what is common and what is different between *straight* and *droit*. The comparison becomes much easier and clearer if the comparison is carried out at the level of the senses. It is then easy to observe that the two words share exactly three senses: *(rectilinear)*, *(directly)*, *(honest)*; that French *droit* adds to these a sense *(right-hand)*, while English *straight* adds a number of other senses which have no equivalent in French.

This configuration may be illustrated visually in the form of two overlapping sets (Figure 1). The elements of the sets are the senses, presented here in no specific order. The sets themselves refer to the lexical units – the words – that happen to group these senses in their own polysemies. One may talk here of two “overlapping polysemies”.



to detect certain typological tendencies among the lexical structures of the world's languages, and eventually pave the way for the formulation of lexical universals.

### 3. Towards a typology of colexification

For each specific notion taken as the object of study (see 4.3.), the empirical method here adopted follows two steps:

- First, select the word that lexifies this notion in one language, and identify the various senses which form part of its polysemy, in this particular language.
- Second, once a list of senses has been proposed for this first language, observe a second language, to see which of these senses are also lexified together (or "colexified"), and what new senses have to be added to the list. Then proceed to another language, and expand the list accordingly.

To use a chemical metaphor, one could say that the comparison of different molecules requires first to identify the nature of the atoms that take part in their structure (3.1.); and then, once each molecule has been broken up into its components, to observe the bonds that connect these atoms together (3.2.).

#### 3.1 Senses: The atoms within each molecule

Imagine we want to observe the various polysemies attested cross-linguistically around the notion "rectilinear". The first step is to select, in any language, a word that may translate ("lexify") this notion; for example, English *straight*. What now has to be done, before being able to compare it with a word from another language – or with another word of the same language – is to break down this lexical unit into its own various senses.

Most of the time, this is done intuitively, as probably most dictionaries do: obvious functional considerations seem sufficient to analyze, say, *(rectilinear)* and *(frank)* as two distinct senses, deserving separate treatment. However, on some occasions one may object to the arbitrariness of such intuitive choices, when two senses appear to be so close, that their distinction might be an artifact of the linguist's analysis. In the case of *straight* (2.2.), for instance, one may argue that the psychological senses *(frank)* and *(honest)* form in fact a single meaning for the native speaker of English, so that we are dealing with a case of vagueness *(frank, honest)* rather than a case of polysemy, strictly speaking, between two separate senses.

Trying to resolve such a tricky debate with a definite answer might result in unverifiable and irreconcilable points of view. Luckily, there is one way out of this dilemma, which is to base all sense distinctions upon the empirical observation of contrasts between languages. For example, the fact that French lexifies *(rectilinear)* with *(honest)* but not with *(frank)* suffices to justify the choice of distinguishing between the two

latter meanings as if they were two separate senses. Even though this may fail to represent faithfully the language-internal perception of an English native speaker, at least this serves efficiently the purpose of cross-linguistic comparison: it becomes then easy to state the facts by saying that these two senses are treated the same in English, and not in French. The same reasoning would apply to *(directly)* and *(immediately)*, which despite their semantic closeness, must be distinguished due to the different treatment they receive in French. The repetition of the same procedure, for each word under scrutiny, makes it possible to define with precision the list of its possible senses.

This empirical method of defining senses based on cross-linguistic comparison has the valuable advantage that it helps "sidestep the vexing problem of distinguishing between polysemy and vagueness" (Haspelmath 2003: 231). Now, a corollary of this approach is that the list of senses for a given word is likely to evolve during the process of cross-linguistic comparison. Indeed, the more languages are considered, the more new distinctions are likely to be found, thereby resulting in the need to split up certain senses that were initially not distinguished. For example, suppose the examination of nine languages showed the meaning *(horizontally rectilinear)* to be always lexified in the same way as *(vertically rectilinear)*: this would result in the initial grouping of these two meanings as a unique vague sense *(rectilinear (horiz. or vertic.))*, with no empirical reason for splitting it in two. But once a tenth language is considered that forces to make this distinction, then the former sense *(rectilinear)* will have to be cracked down into two separate senses, for the purpose of cross-linguistic comparison. As a result, the description given for each polysemous lexeme in the first nine languages may have to be revised, due to the introduction of a new semantic distinction after the tenth language has been examined.

Note that this remark is not necessarily an issue for the semantic analysis itself: one will simply have to describe *(horizontally rectilinear)* and *(vertically rectilinear)* as two potentially separate senses, which simply happen to be formally indistinct in the first nine languages, but distinguished in the tenth. The problem rather arises at the practical level, if one thinks of setting up a typological database: for it means that the semantic descriptions made at a given point in time, during the constitution of the database, are likely to evolve as more and more distinctions are considered from new languages. This can entail the necessity for the first languages entered in the database to be reassessed again and again as the list of descriptive senses grows. When this takes the form of a semantic map (section 4), this also means our maps will have to integrate the capacity to evolve constantly, and adapt to whatever new input comes in. This is probably feasible, but likely to raise certain technical questions.<sup>3</sup>

3. In the grammatical domain, Haspelmath (2003: 231) reassures us on this point, by saying: "the typical experience is that after a dozen languages have been examined, fewer and fewer

Despite these potential issues on the practical side, it is important to see that this method, by basing every semantic distinction on empirical data, provides a safe antidote against the vagaries of intuition; it ensures that the whole process of semantic analysis is always verifiable – and therefore falsifiable.

### 3.2 Colexification: The bonds between the atoms

In itself, the result of the preceding step pretends to be little more than a list of notions (senses). For one thing, these notions can be shown – using the cross-linguistic method described above – to be functionally distinct from each other; but at the same time, the way they were compiled implies that they are potentially linked together in at least some of the world's lexicons. No particular claim is being made at this stage, except that this non-arbitrary selection of notions should provide a useful “etic grid” against which language-specific, “emic” categorizations are to be observed.

But what is really relevant to our typological study is not so much these atoms *per se*, as the bonds that each particular language creates between them. Once a list of senses is arrived at, the phenomenon most relevant for the second stage of observation may be called COLEXIFICATION.

- (1) A given language is said to COLEXIFY two functionally distinct senses if, and only if, it can associate them with the same lexical form<sup>4</sup>

For example, Figure 1 showed that English colexifies the senses ⟨immediately⟩ and ⟨undiluted⟩; ⟨rectilinear⟩ and ⟨right-hand⟩ are colexified in French; ⟨rectilinear⟩ and ⟨directly⟩ are colexified both in English and in French. One of the advantages of the term “colexification”, which I am proposing here, is to be purely descriptive, and neutral with respect to semantic or historical interpretations – contrary to the term “semantic shift”, chosen for example by Anna Zalizniak (this volume).

One interest of the colexification model is to be readily exploitable for typological research. For example, one may want to check what proportion of the world's languages colexify the two senses ⟨rectilinear⟩ and ⟨honest⟩, as French and English do: is this connection found only in a few scattered languages? Is it an areal phenomenon covering, say, Western Europe? Is it well represented in other parts of the world? Or is it universally common?

---

functions need to be added to the map with each new language.” It remains to be seen whether this comforting statement also applies to the richer realm of lexicons.

4. The term “lexical form” may refer to a lexeme or a construction, or occasionally to a lexical root (but see below for a discussion).

Incidentally, because the list of senses is initially based on the polysemy of a specific word in a given language, it is logical that the first stage of the observation will show these senses to be colexified in the language under consideration. For example, because the initial sense list was built as the description of English *straight*, then it necessarily results that these senses are all “colexified” in English. At this stage of the research, due to a bias in favour of the language taken as the starting point, such an observation is circular, and has little interest. But these representations become rapidly more informative as other languages are considered. For instance, French adds to the list a new sense ⟨right-hand⟩, which is not lexified by English *straight*. As more languages are explored, and the list aggregates more and more senses, it will eventually come closer to a universal grid of potentially interconnected notions – with less and less risk of an ethnocentric bias in favour of a specific language.

### 3.3 Strict vs. loose colexification

Strictly speaking, the notion of colexification should be understood as “the capacity, for two senses, to be lexified by the same lexeme in synchrony”. However, nothing prevents the model from being extended, so as to make provision for several – hierarchized? – levels of colexification. These may include the linking of two senses by a single lexeme across different periods of its semantic history (e.g., *droit* also meant “right, true” in Old French); their association in the form of doublets (e.g., Fr. *droit* and *direct*), or other etymologically related forms (Eng. *straight* and *stretch*); the impact of lexical derivation (Eng. *straight* → *straighten*; Fr. *droit* → *droiture* “honesty”) or composition (Eng. *straight* → *straightforward*); and so on.

Ideally, for the sake of accuracy and future reference, the different types of formal relations should be kept distinct in the representation of the data, e.g., with the use of different symbols. In particular, “strict colexification” (same lexeme in synchrony) should be carefully distinguished from “loose colexification” (covering all other cases mentioned here). This will be done here formally, in tables (section 5.2., and Appendix 2), with the use of respectively “+” vs “[+]” signs; and in maps (section 7.3.), with the use of solid vs. dotted lines. To take an example, one can represent the colexification of ⟨rectilinear⟩ and ⟨honest⟩ in English as “strict colexification”, because both can be lexified with exactly the same form in synchrony (*straight*). As for the sense ⟨simple, easy to understand⟩, it can also be said to be somehow part of the lexical field of *straight*, but only indirectly, through the compound form *straightforward*; in other words, English shows “strict colexification” between ⟨rectilinear⟩ and ⟨honest⟩, but “loose colexification” between ⟨rectilinear⟩ and ⟨simple⟩.

Finally, in the framework of a typological survey carried out around a specific notion, I propose that the senses to be included in the universal list – and in the map derived from it – should fill one condition: that is, they should only include those

senses that are attested to be in *strict colexification* in at least one language of the world. For example, supposing one language  $L_1$  were found where exactly the same form in synchrony might translate both senses ⟨rectilinear⟩ and ⟨simple⟩, then this would be a sufficient condition for the latter sense to be included in the sense list associated with ⟨rectilinear⟩. This being done, it will be possible to state that some languages – like  $L_1$  – colexify these two senses directly (“strictly”), while others – like English – colexify them only indirectly (“loosely”), and others again do not colexify them at all. Conversely, if no language can be found where the two senses are strictly colexified, then it is probably a safe principle to exclude them from the sense list, to avoid the risk of widening and blurring indefinitely the boundaries of a polysemous network. This principle will be useful, for example, in 5.3.2., when discussing the relationship between “breathe” (Latin *spīro*) and “die” (Latin *ex-spīro*). Indeed, because these two senses often show some specific semantic relationship through lexical derivation, it would be tempting to include them in the same sense list, and consider them as indirectly colexified. However, because no language can be found – for obvious reasons – where these two senses are expressed by exactly the same form (“strict colexification”), it is preferable that the sense “die” be kept away from the sense list of “breathe”.

### 3.4 Interpreting colexification

In principle, the colexification model itself consists first and foremost in stating the facts – that is, detecting and documenting the cases of colexification that are empirically attested across languages. The interpretation of these semantic connections, whether it takes a historical or a cognitive perspective or otherwise, arguably belongs to another phase of the study.

For each pair of senses  $s_1$  and  $s_2$ , several configurations may come out of the data, suggesting possible questions for the typological study of the lexicon.

In case the colexification of  $s_1$  and  $s_2$  appears to be attested nowhere, this may be because the two senses are directly opposite – e.g., ⟨rectilinear⟩ vs. ⟨curved⟩; cognitively divergent – ⟨rectilinear⟩ vs. ⟨slow⟩; or simply unlikely to be related – ⟨rectilinear⟩ vs. ⟨green⟩.

If two senses  $s_1$  and  $s_2$  are colexified in at least one language, this is normally the sign – setting aside the case of accidental homophony – that the human brain has proven able to perceive these senses as somehow “semantically connected”. This connection may be direct or indirect, via historical paths that may or may not be still perceived in synchrony. It is then the purpose of semantic or etymological studies, to propose a convincing explanation for that connection: is the relationship between  $s_1$  and  $s_2$  a case of metaphor, metonymy, hyperonymy, analogical extension ...? Is it possible to reconstruct the direction taken historically by this extension (from  $s_1$  to  $s_2$ ,

or the reverse)? Is it useful to reconstruct a missing link<sup>5</sup> between two senses whose relation is intuitively opaque?

Sometimes one may want to take the reverse perspective, and try to answer the question why a language does *not* colexify two senses  $s_1$  and  $s_2$ , that is, treats them separately, when other languages treat them alike. Most often, this state of affairs will be simply considered, just like many other linguistic features, to result from a chance distribution between languages. In some cases, however, hypotheses may be proposed that would draw a correlation between a specific case of colexification (or of non-colexification), and, say, the language’s environment. For example, Brown (2005a) suggests that the colexification of ⟨hand⟩ – ⟨arm⟩ may be influenced by the geographical situation of the community. According to him, the use of “tailored clothing covering the arm” in colder environments tends to make the contrast between the hand and the arm more salient, thus favoring the existence of two separate lexical items. Likewise, Brown (2005b) sees another correlation between the lexical distinction ⟨finger⟩ – ⟨hand⟩ and cultural practises in terms of farmers vs. hunter-gatherers.<sup>6</sup> Regardless of the likelihood of these hypotheses, it is instructive to see that the facts of colexification may receive various sorts of functional explanations, whether semantic, historical, cognitive or cultural – thereby opening fascinating debates.

It may be a subject for discussion, how one should interpret the statistics of colexification. That is, supposing the colexification of  $s_1$  and  $s_2$  is particularly widespread in the world’s languages, should we see this as a sign that these two senses are particularly “close”? that their semantic connection is – functionally or cognitively – particularly “tight”? This brings in the intuitively appealing notion of degrees of “closeness” in the semantic connection. For example, supposing ⟨rectilinear⟩ and ⟨honest⟩ turned out to be statistically more often colexified than, say, ⟨rectilinear⟩ and ⟨right-hand⟩, one may think that the first pair of senses is more deeply motivated than the second pair (?). Admittedly, however, it may be debatable whether semantic closeness should be measured, as I am tentatively proposing here, on statistics based on actual colexification data – rather than assessed, say, on the basis of each notion’s ontological properties.

5. For example, the spatial notion ⟨rectilinear⟩ is metaphorically associated with social normality (cf. *the straight and narrow*), as opposed to eccentricity or originality; hence such senses as ⟨classical⟩, ⟨not homosexual⟩, ⟨not on drugs⟩, etc. In this case, the meaning ⟨satisfying the social norm⟩ could be described as the missing semantic link – whether in diachrony or in synchrony – between several members of this polysemous network.

6. “Languages of farmers tend more strongly to lexically distinguish “finger” from “hand” than those of hunter-gatherers, which tend more strongly to use a single term to denote both “finger” and “hand”” (Brown 2005b: 527). Brown’s rather unconvincing hypothesis resorts to the saliency of the finger in those societies which make use of finger rings; he claims that this cultural habit is more developed among farmers.

Crucially, the descriptive concept of colexification lends itself to just the same sort of observations, tests and representations as any other language feature. For example, specific pairings of senses may be represented in the form of geographical maps, using isoglosses or coloured spots.<sup>7</sup> Instances of colexification may be attributed to genetic subgroups and protolanguages (see François forthcoming), or result from local innovations. They may also be borrowed through language contact, and take part in areal phenomena. They may change through time, be subject to analogical levelling, and so on and so forth.

To take just one example, the senses ⟨hear⟩ and ⟨feel⟩ are colexified in several areas of the world: Catalan *sentir*, Italian *sentire*, Mwotlap *yoñteg*, Bislama *harem*:

- Knowing that Latin lexified distinctively *sentire* “feel” and *audire* “hear”, Catalan and Italian evidently illustrate a case of late semantic merger between the two words. Historically speaking, this is a parallel innovation in these two languages, whether due to areal or to typological convergence.
- Conversely, the colexification ⟨hear⟩ – ⟨feel⟩ found in Mwotlap is also attested in all known languages of Vanuatu, and was demonstrably inherited from a similar pattern in their common proto-language: Proto North-Central Vanuatu \*roño “hear, smell, feel” (Clark n.d.), from Proto Oceanic \*ronjor.
- Finally, the presence of exactly the same colexification in Bislama, the pidgin/creole of Vanuatu, historically results from language contact. The verb *harem* “hear, smell, feel”, despite reflecting English *hear him* in its form, borrows its semantics directly from the lexical structures of Oceanic languages, the vernacular substrate of Bislama (Camden 1979: 55–56).

In sum, colexification may result historically from typological convergence, from genetic inheritance, or from contact-induced change ... just like any other structural feature of a language.

### 3.5 Typological prospects

The observation of colexification does not only provide insights on individual languages or language groups. In theory, one can also conceive the possibility of formulating typological hypotheses in this domain, just like in other domains of language research. The following paragraphs attempt not to state actual facts – most examples in this section being hypothetical – but to define the form that future research will be able

7. See Brown's maps on the colexification of “hand”/“arm”, “finger”/“hand”, as well as and Kay & Maffi's on “green”/“blue” or “red”/“yellow”, in the *World Atlas of Language Structures* by Haspelmath et al. (2005).

to give to the formulation of universals, whether absolute or implicational, regarding the typology of colexification.

An absolute universal would take the form (2), or its shorter equivalent (2'):

- (2) ‘If a language lexifies sense s1 with the form X, then sense s2 will be lexified in the same way.’
- (2') ‘All languages colexify the pair of senses s1–s2.’

An example of this – of course subject to empirical check – could be the pair ⟨male fish⟩ vs. ⟨female fish⟩, which is apparently never formally split in the world's lexicons – as opposed to mammals, for which separate lexification is common. Interestingly, this theoretical case should normally not come up in the data, given the method chosen to distinguish between senses in the first place (3.1.): the condition was to retain only those sense distinctions that are attested in at least one language. Yet a lighter version of (2'), in terms of statistical tendencies (“Most languages colexify...”) would be perfectly acceptable, as would its symmetrical counterpart (“Very few languages colexify ...”).

As for implicational universals, they can associate two cases of colexification:

- (3) ‘If a language colexifies s1 and s2, then it will also colexify s3 and s4.’

For example, a likely assumption would suggest that if a language colexifies ⟨arm⟩ and ⟨hand⟩, then it will do the same for ⟨leg⟩ and ⟨foot⟩.<sup>8</sup> Or if it colexifies ⟨paternal uncle⟩ and ⟨maternal uncle⟩, then it will also colexify ⟨paternal aunt⟩ and ⟨maternal aunt⟩ ...

A subtype of this formula would be (3'):

- (3') ‘If a language colexifies s1 and s2, then it will also colexify s2 and s3.’

For example, if a language colexifies ⟨upper arm⟩ and ⟨hand⟩, then it will probably colexify ⟨forearm⟩ and ⟨hand⟩ too. Or, if ⟨word⟩ and ⟨language⟩ are colexified, then ⟨speech⟩ should be able to take the same form. As these (fictitious) examples suggest, this sort of formula typically applies when the three senses can be conceived as showing some form of – logical, cognitive ... – ordering, so that s3 typically comes “between” s1 and s2: e.g., because the forearm is physically located between the upper arm and the hand, the colexification of the latter two makes it likely that the item in the middle should be lexified identically. The case of *word* < *speech* < *language* which I intuitively suggest here would illustrate a similar, but more figurative, case of ontological hierarchy between referents.

8. Counterexamples to this potential universal can however be found, such as Lo-Toga (Torres Is., Vanuatu, Oceanic group; pers. data), where ⟨arm⟩ = ⟨hand⟩, but ⟨leg⟩ ≠ ⟨foot⟩.

Another variant of this formula would be (4):

- (4) ‘If a language colexifies  $s_1$  and  $s_2$ , then it will not colexify  $s_2$  and  $s_3$ .’

– which may be also formulated as:

- (4') ‘Although there may exist languages that colexify  $s_1$  and  $s_2$ , and others that colexify  $s_2$  and  $s_3$ , no language colexifies together the three senses  $s_1-s_2-s_3$ .’

One possible example of such a formula would be: if a language colexifies ⟨person⟩ and ⟨male person⟩, then it will not colexify – that is, it will treat distinctively – ⟨male person⟩ and ⟨husband⟩. Indeed, one can think of many languages where ⟨person⟩ and ⟨male person⟩ go together (as with French *homme*), and many languages where ⟨male person⟩ is the same as ⟨husband⟩ (as with Latin *vir*), but none – until further research is done – where the three are lexified the same.

Another kind of implicational universal would associate colexification with a criterion outside the lexicon, as in (5)–(5'):

- (5) ‘If a language colexifies  $s_1$  and  $s_2$ , then it will have the linguistic property P.’  
 (5') ‘If a language has the linguistic property P, then it will colexify  $s_1$  and  $s_2$ .’

An example of (5) could be a statement about parts of speech, such as: if a language colexifies ⟨black⟩ and ⟨darken⟩, then it treats adjectives as (a subclass of) verbs. A possible illustration of (5') would be something like: if a language doesn't distinguish count nouns from mass nouns, then it will colexify ⟨wood⟩ and ⟨tree⟩. Needless to say, all these examples are intuitive, and would only make sense if confirmed by relevant empirical data.

Finally, one could conceive possible correlations between certain instances of lexification and specific properties of the language's environment. This would lead to universals – or at least tendencies – such as:

- (6) ‘If a language colexifies  $s_1$  and  $s_2$ , then its environment will have the property P.’  
 (6') ‘If a language's environment has the property P, then this language will colexify  $s_1$  and  $s_2$ .’

The term “environment”, used in a functional perspective, encompasses all properties that are not strictly linguistic, but which are somehow associated with the language or its speaking community. One could thus imagine the following sort of hypothesis: if a language colexifies ⟨cow⟩ and ⟨bull⟩ under a single term, it is likely that this language is used in a society where this particular gender difference is functionally less relevant – that is, where cattle farming is not practiced traditionally. Similar types of correlation were mentioned in 3.4. above, with Brown's cultural-cognitive interpretations of certain cases of colexification.

#### 4. Lexical semantic maps

So far, the method here exposed has entailed the identification and manipulation of senses, in the form of unordered lists. The only visual representation proposed (Figure 1) took the simple form of overlapping sets, still with no specific internal organization. Yet the high number of senses involved, combined with the number of languages potentially explored, would ideally require defining a more sophisticated way of organizing and presenting the results of our semantic observations. This is what I will now propose to do, in a form suggested by current research in grammar typology,<sup>9</sup> and systematized by Haspelmath (2003): SEMANTIC MAPS.

##### 4.1 General principles of semantic maps

Here is how Haspelmath (2003: 213) defines semantic maps: “A semantic map is a geometrical representation of functions in “conceptual/semantic space” that are linked by connecting lines and thus constitute a network.”

Essentially speaking, a semantic map takes the form of a two-dimensional chart, and represents a selection of meanings (“senses” in my terminology, “functions” in Haspelmath's). These meanings are ordered in space according to certain principles, and explicitly interconnected, thus forming a semantic network. In itself, this semantic map constitutes an etic grid which claims to be language-independent, “a coherent chunk of a universal network”. This universal grid then serves to visualize the “emic” categorizations which are made by each specific language: for a given form in a given language – usually understood in synchronical terms – it then becomes possible to identify, on the universal map, those meanings that are covered by this form, and those that fall without its scope.<sup>10</sup>

The whole methodology presented by Haspelmath is compatible with the model of lexical typology which I here propose to develop. The only difference is that he explicitly designs his model as a way to represent “the geometry of grammatical meaning”, while the present discussion deals with the lexicon. Yet, even if all his examples are taken from facts of grammar, he himself suggests that his model should theoretically be compatible with the lexical domain too (2003: 237). In a way, the following pages may be seen as an attempt to apply to the lexicon the principles defined by Haspelmath for drawing semantic maps.

9. See, for example, Anderson (1982) for the perfect; Croft et al. (1987) for the middle voice; Jurafsky (1996) for the diminutive; etc.

10. For a visual illustration of this principle, see Figure 4, and the figures in Appendix 3.

Of course, several authors have already proposed to represent lexical semantics, and in particular polysemous networks, in the visual form of a diagram or map. But most often, their intention was to illustrate a pattern of polysemy specific to one language, or one group of languages.<sup>11</sup> By contrast, the maps I propose to draw here claim to have a universal value, that is, to provide results that are virtually independent from any particular group of languages. This is coherent with the stance taken by Haspelmath for his grammatical maps: “The configuration of functions shown by the map is claimed to be universal” (2003: 217). Of course, the quality and precision of a map will depend on the number and genetic diversity of the languages observed. But essentially, whatever result comes out of such a study, should be able to claim universal relevance. An important consequence of this principle is that any new data from a natural language should therefore be able to falsify these results. As Haspelmath (2003: 232) puts it, “Every semantic map can be interpreted as making a universal claim about languages, that can be falsified easily.”

#### 4.2 Connecting senses together

In comparison with the simple format of a sense list, the main interest of semantic maps is to organize the polysemous network in a way that makes explicit the various semantic connections between these senses. This is shown visually, on the one hand, by the iconic grouping of close senses in contiguous areas of the map; and on the other hand, by the use of explicit connecting lines to visualize semantic paths.

Judgments of closeness between senses are established in a dual fashion: first, by taking into account the ontological properties of each sense; second, by examining empirical data from various languages.

The intrinsic ontological properties of each sense can legitimately be taken into account in order to suggest a semantically plausible ordering between senses. For example, suppose one came across an array of senses such as the one observed with the verb *ōl* in Mwotlap (François, in prep.):<sup>12</sup> {creak}; {name a child so-and-so}; {crow};

11. Thus, the maps found in Matisoff (1978) intend to represent certain semantic associations specific to the Tibeto-Burman family; those in Evans (1992) or Evans & Wilkins (2000: 560) apply to Australian languages; Enfield (2003) to Southeast Asian languages; Tyler & Evans (2003 [2001]: 125) propose a semantic network specific to the polysemy of English over ... Some projects aim at representing semantic associations at the level of the whole lexicon, but they are still, by definition, restricted to a single language – cf. Gaume et al. (this volume) for French; or software such as Thinkmap’s *Visual Thesaurus*<sup>®</sup> for English. ([www.visualthesaurus.com](http://www.visualthesaurus.com))

12. Even though this list of senses, as well as its representation in Figure 2, are drawn after the polysemy of just one word in one language, I propose that it is fictitiously understood, for the purpose of this demonstration, as if resulting from cross-linguistic comparison. Indeed the forms of reasoning that apply in both cases – whether we consider one polysemous network,

{mention s.o.’s name}; {yell}; {invoke a divinity}; {scream}; {bark}; {hail s.o.}; {call s.th. such-and-such}... This kind of simple list, presented in random order, makes it hard to identify the semantic links between these senses. But functional considerations allow certain senses to be grouped according to their common semantic properties. Thus, several senses refer to the emission of intense high-pitched sounds, whether by humans ({yell}, {scream}), animals ({crow}, {bark} ...) or objects ({creak}). Other senses refer to human social activities that consist in uttering the name of another person; this can be done for the purpose of calling out to someone ({hail s.o.}, {invoke a divinity}), or for the purpose of referring to them ({mention s.o.}). Finally, the act of uttering a name may refer to the social act of giving a name to someone – typically a child – or to something.

Semantic connections can then be proposed, which chain senses according to their functional similarities. These connections may then easily be represented in space, in the form a visual graph such as Figure 2:

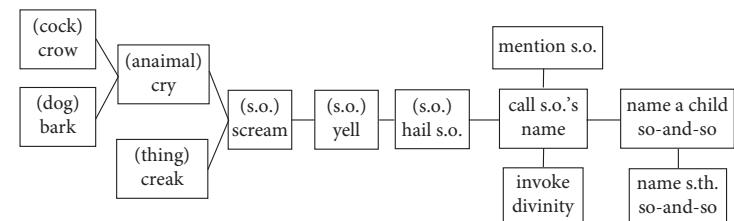


Figure 2. Senses may be linked based on functional properties.

Crucially, because the semantic connections here proposed are supposedly based on *ontological* properties of the notions referred to, this means they must normally be conceived as independent of any specific language. That is, even though the list of senses itself was initially based on the observation of actual languages, ultimately the fact that a sense *s2* will be understood as forming the missing semantic link between *s1* and *s3*, should not depend on any particular language, but simply on the intrinsic properties of each sense. For example, the act of “hailing someone (by shouting their name)” constitutes a logical transition between “shouting (in general)” and “uttering s.o.’s name”. This organization of meaning must be understood as driven not by idiosyncrasies of any specific language, but rather by universal characteristics of the real

or the intertwining of several such networks into one – are fundamentally the same, at this particular stage of the study.

world – or more exactly, of the world as it is perceived by the human brain and filtered by human activities.

This being said, it remains obvious that the connections proposed between each two senses, and more generally the semantic map that results from these connections, cannot be directly observed in the material world, and thus constitute *hypotheses* on part of the observer. This means that they must be amenable to proof or demonstration, that they are open to debate, and that they should be falsifiable. The problem is, at least some of these semantic hypotheses – about how two senses should be connected – may ultimately depend on the linguist's intuition. This is potentially an issue (see also the discussion in 3.1.), because the conscious representations of the world by an observer do not necessarily match the subconscious connections which are actually made by the speaker's brain. It is therefore necessary to define a method of falsification that would rest on empirical observation.

The method suggested by Haspelmath indeed resorts to observable data from actual languages. The basic idea is that senses should be arranged in space in such a way that each lexical unit in one language “occupies a contiguous area on the semantic map” (2003: 216). Furthermore, each specific connecting line should reflect the existence of at least one attested case of a direct lexical connection between these two senses, in any of the world's languages. Thus, supposing one language were found that only colexified a sense  $s_i$  and a distant sense  $s_j$ , but none of the other senses tentatively proposed in-between, then the background map should be redesigned, and a “shortcut” connecting line added between these two senses. Conversely, if all words colexifying  $s_i$  and  $s_j$  also include, in their polysemy, the various intermediate steps proposed along the functionally-based semantic chain, then the hypothetical map can be said to be confirmed by empirical data.

Incidentally, it may happen, on some occasions, that two distinct paths may be defined in order to relate two senses on the map, with no strong reason for choosing between these two paths. We shall see precisely an example of this in 5.3.4., where two different semantic hypotheses will be shown to equally account for the colexification of *⟨breath⟩* and *⟨supernatural power⟩*. Insofar as this sort of hypothesis is also supported by empirical data – in this case, the existence of two distinct sense chains attested in the world's languages – nothing prevents us from representing this double path on the map.

#### 4.3 Choosing a pivot notion

In section 3 above, I briefly mentioned the necessity to choose a specific notion (sense) as the pivot of the map. This requires justification, especially because this principle seems to differ from Haspelmath's (2003: 232) method for drawing grammatical maps.

Choosing a specific sense (e.g., *⟨hail s.o.⟩*) as the pivot entails that the empirical data to be observed must consist exclusively of lexical units that specifically include

this sense in their polysemy. This important requirement is a precaution against the risk of starting an open-ended map with ever-shifting boundaries. For example, consider the colexification of *⟨hail⟩* and *⟨(animal) cry⟩*. If *⟨hail⟩* were not given any special status, nothing would then prevent us from including in the data words that encompass *⟨(animal) cry⟩* as one of their senses, yet having no connection with *⟨hail⟩*: for example, a verb meaning “(animal) cry; (s.o.) cry out in pain; weep ...” (cf. Eng. *cry*). If this were allowed, then the map would extend so as to include all the semantic connections associated with the sense *⟨weep⟩*, and so on and so forth. Such a map with no center would shift indefinitely ... so as to gradually enclose the whole lexicon. Despite the immense interest of potentially achieving a map that would represent the global “geography of the human mind” (Croft 2001), such a configuration would rapidly lead to uncontrollable results that would raise obvious technical issues, and whose significance in terms of scientific information would end up being questionable. It is therefore safer to circumscribe in advance the scope of the map that is to be drawn, by providing one sense with the special status of pivot or centre. Incidentally, I propose to translate typographically the special status of the pivot notion, by using small uppercase and braces – e.g., {HAIL} – as opposed to the other senses of the network – e.g., ⟨bark⟩.

There is a corollary to this principle. If two senses  $s_1$  and  $s_2$  are attested to be colexified in the world's languages, the map centered on  $s_1$  will be a different map from the one centered on  $s_2$ . Thus, the choice of {HAIL} as the pivot will trigger a specific semantic network – one that can be called, in short, the “lexical map of {HAIL}” – which will tell a totally different story from the choice of {(animal) CRY}. Quite logically, however, one can predict that these two maps will have a whole chunk in common – that is, the connection between these two senses, plus whatever further senses are attested to colexify with these two senses together. Thus, supposing a language were found that colexified *⟨(animal) cry⟩* – *⟨scream⟩* – *⟨hail⟩* – *⟨call⟩*, then each of the four semantic maps centered on each of these senses would necessarily have to include this particular chain of senses – along with other ramifications specific to each map.

Finally, note that the status of pivot of a lexical map has nothing to do with the notion of prototype, which is only relevant to the description of individual lexemes. Thus, it is perfectly possible that a typological map centered on the sense {HAIL} incorporates a lexeme X whose polysemy encompasses only those senses that appear to the left of *⟨hail⟩* in Figure 2 above (*⟨hail⟩*, *⟨scream⟩*, *⟨creak⟩*, *⟨(animal) cry⟩* ...). In this particular language, it is likely that a prototype-based approach would describe this word X as being built around the prototypical meaning “shout with high-pitched voice, scream”; the sense *⟨hail⟩* would be nothing more than a peripheral offshoot of that core meaning – regardless whether or not it is the pivot of the universal map that includes it.

Another difference is that the definition of a prototypical meaning, in the (language-internal) description of a word, constitutes an interpretative claim about

this word that may be challenged or falsified. On the contrary, the selection of a given notion as the pivot of a (universal) lexical map entails no claim at all: it is simply an arbitrary choice, the starting point before any lexical map may even begin to be drawn.

### 5. Elaborating a universal map for “breathe”

In order to illustrate in full detail the typological method I am here advocating, I now propose to delve into a specific notion, and build the lexical semantic map that will best render the various polysemies associated with it in the world’s languages. This is what I will do in the remainder of this article, around the notion “BREATHE”.

The notion “breathe” is here understood as the physiological activity of breathing characteristic of humans and animals. I will first observe, for each language of the corpus, the set of other senses with which this notion is colexified. Then I will attempt to draw the lexical map of the notion {BREATHE}. The final form taken by these two steps appear respectively as Table 2 and Figure 5 in Appendix 2.

This small case study rests on a corpus of 16 lexical headwords in 13 genetically diverse languages. Each entry consists of either a single word, or a lexical root, in which case several words are encompassed under the same entry. In particular, it is frequent that the noun and the verb associated with the notion {BREATHE} differ formally from each other; in this case, I have organized arbitrarily the data in the appendices in such a way that the default headword is the noun, while the cognate verb, when formally different from it, has a secondary status (loose colexification).<sup>13</sup>

The lexical database presented in Appendix 1 shows a total of 114 words involved in the comparison. Of course, richer data, taken from more languages, would logically result in richer results, with even higher typological significance. However, the corpus here analyzed was judged at least sufficient for the purpose of illustrating the typological method here proposed.

#### 5.1 A first overview of the verb “breathe”

In Makonde, a Bantu language of Tanzania, the verb *ku-pumula* colexifies {breathe} and {take a rest}. This semantic connection has a transparent motivation. In the first place, the physiological act of breathing becomes particularly significant – “cognitively salient” – after one has held his breath while making a physical effort. The act of sitting

<sup>13.</sup> This is why Table 2 shows plain “+” signs in the rows {act of breathing} and {puff of breath}, but bracketed “[+]” signs, standing for loose colexification, in the first row {breathe}. See also the isolectic sets in the maps of Appendix 3.

down for a minute after an intense effort, or even of resting for a whole day after a week of work, can be seen as a semantic expansion of this initial meaning, even when what is relevant is not so much the act of breathing *per se*, as that of ceasing an effort. If we add {pause for breath} as the missing semantic link (3.4.) between these two senses, the polysemy of *ku-pumula* can be represented using a string of three senses {breathe} – {pause for breath} – {take a rest}. This is a classical case of colexification originating in semantic extension.

Makonde is not the only language to have developed this polysemy. English shows a case of loose colexification (3.2.) between *breathe* and *take a breather*. The colexification, whether strict or loose, of {breathe} and {take a rest} is also attested in Sar (noun *koo*), in Arabic (root *r.w.h*), in Nahuatl (verb *imi’iyo*), in Mwotlap (verb *mökheg*), in Nélémwa (root *horêâ-*), in Russian (root \**du[x]*); but not in Latin, Greek or Inuit. This is enough evidence to propose this case of colexification as typologically significant.

Interestingly, Mwotlap *mökheg* can also equally be used for any period of rest, i.e., not only minutes of pause within hours of work, but also days of pause within months of work – that is, what we would call “take a vacation”. Since certain languages do not go that far in the semantic expansion of {BREATHE} (e.g., English would hardly describe a month-long holiday as *take a breather*), it is wiser to define formally not three but four different senses here: {breathe}, {pause for breath}, {take a rest} and {take a vacation}. Out of these four senses, we will say that English colexifies only three, whereas Mwotlap covers them all. Incidentally, this proposal does not involve the claim that these senses are necessarily distinct for the Mwotlap speaker – and it is perfectly likely that {take a rest} and {take a vacation} should be grouped together under an emic approach. But what is relevant here, for the specific purpose of language comparison, is that these two functional situations are colexified in Mwotlap, but distinguished in English; hence the choice to treat them, in an etic perspective, as if they were distinct semantic units (see discussion in 3.1.). Incidentally, Russian *otdyx*, etymologically connected with *dyšat’* “breathe”, means both “rest” and “vacation”.

In a similar way, the Nélémwa verb *horêâñ* has added an extension to the meaning {take a rest}, namely {stop doing s.th., cease} (e.g., *Co horêâñ o khiboxa pwaxim tavia* “Stop beating your dog!”). This semantic offshoot clearly adds a new sense to the potential polysemy of {BREATHE}.

The same observations can be made for other senses related to {BREATHE}. In some languages (e.g., Greek *pneō*), the same verb is used for {breathe}, for {blow} (i.e., a person blowing actively into s.th., like a flute) and/or for {(wind) blow}. A further connection that is sometimes attested is between {blow} and {whisper}, with a shift towards the notion of articulated speech. Thus in Araki (François 2002), the verb *soro* connects the notions {blow, puff}, {blow into s.th.} and {talk, tell a story} – see also the derived noun *sorosoro* “speech, story; language”. Likewise, the French verb *souffler* means both {blow, puff} and {whisper, prompt}.

But the latter example of colexification potentially raises an issue, because it involves the sense ⟨blow⟩ rather than ⟨breathe⟩ (“breathe” is *mapu* in Araki, *respirer* in French). Consequently, it should be kept aside from the semantic network of {BREATHE} strictly speaking, to avoid the risk of shifting the center of observation from one sense to the other, and thus expanding infinitely each polysemous network (see 4.3.). In other words, the evidence so far allows us to include ⟨blow⟩ among the senses directly connected to {BREATHE}, and ⟨utter⟩ among the senses directly connected to {BLOW}; but it does not illustrate any colexification between {BREATHE} and ⟨utter⟩.

Of course, the conclusion would be different if we came across languages that did witness the colexification of these two senses. This is in fact the case with the noun *horêâ-* in Nélémwa, which means both ⟨breath, breathing⟩ and ⟨spoken message⟩. The English phrase *I won't breathe a word* also illustrates the potential connection between ⟨breathe⟩ and ⟨utter⟩, arguably via a missing link ⟨whisper⟩ (as in *breathe a prayer*). These two examples finally legitimize the inclusion of ⟨utter, speak⟩ in the map of {BREATHE}.

## 5.2 From the sense list to the map

Before going any further, it may be useful to recapitulate our first findings in a visual form. A simple way to do so would be to draw a table, based on the list of senses that have been observed to potentially colexify with the pivot notion {BREATHE}. Each column corresponds to one of the languages I have been reviewing so far, representing a subset of my corpus. This leads to Table 1, a partial representation of the sense list under construction here (see Appendix 2 for the complete table).<sup>14</sup>

Table 1. Examples of colexification associated with {BREATHE}

	ENGLISH <i>breathe</i>	RUSSIAN <i>du[x]</i>	MWOTLAP <i>môkheg</i>	NÈLÈMWWA <i>horêân</i>	ARAKI <i>soro</i>	FRENCH <i>souffler</i>
BREATHE	+	+	+	+		
take a rest	[+]	[+]	+	+		+
be on vacation		[+]	+			
cease to do				+		
(wind) blow	+	[+]				+
(s.o.) blow	+	[+]	+		+	+
whisper	+				+	+
utter, speak	+		[+]		+	+

14. The typographical contrast between plain plus “+” and bracketed plus “[+]” corresponds respectively to strict and loose colexification (see 3.2.).

Note that Araki *soro* and French *souffler* are included here for the sake of cross-linguistic comparison. However, as discussed above, they cannot take part in the corpus, because the sense {BREATHE} chosen as this study’s pivot (first row) does not belong to their polysemy.

This representation in the form of a table has the advantage of being clear and straightforward. Yet, it has the drawback of treating all senses on the same level. It may be more interesting to underline the semantic links that relate certain senses with others, and which form functional subsets within the network (see 4.2.). For example, we have seen that the sense ⟨be on vacation⟩ is a semantic extension of the sense ⟨take a rest⟩, itself being closer to the more literal meaning ⟨pause for breath⟩; and that ⟨cease to do⟩ is another, independent offshoot of ⟨make a pause⟩. The chain ⟨breathe⟩ – ⟨pause for breath⟩ – ⟨take a rest⟩ – ⟨be on vacation⟩ thus has a coherence of its own, which is clearly distinct from the chain ⟨breathe⟩ – ⟨blow⟩ – ⟨whisper⟩ – ⟨speak⟩.

A more informative and graphic representation would thus take the form of a semantic map, a diagram showing all the senses attested, together with the most likely semantic connections that link them. These connections are first based on intrinsic semantic properties, and are then checked against empirical data (see 4.2.).<sup>15</sup> This brings about the tentative map of Figure 3.

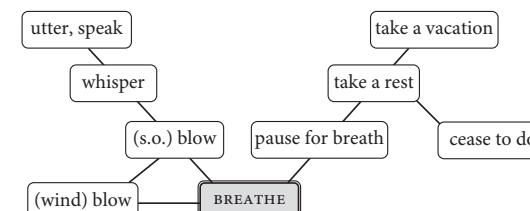


Figure 3. A first semantic map for {BREATHE}.

15. To be precise, the polysemy of Nélémwa *horêâ-* “breath, breathing; spoken message” raises an issue, because it does not include the senses ⟨⟨blow⟩ and ⟨whisper⟩⟩ which functional considerations suggest to posit as intermediate between ⟨breathe⟩ and ⟨speak⟩ (“whisper” in Nélémwa is *nyomamat*). In theory, a rigorous application of the principles exposed in 4.2. should trigger a shortcut line between these two senses. However, the strong functional motivation of ⟨whisper⟩ as a likely missing link, and the fact that the whole chain is empirically attested in other languages, suggests we may be dealing with a case I have not discussed yet: that is, the possibility that an initial chain of senses s1–s2–s3–s4 may have evolved historically so that some intermediate links got lost – via lexical replacement – and only s1 and s4 remained colexified. Although this is debatable, I choose to infringe the rule here, and to keep on the map the intermediate steps of the path, based on functional motivations. This is why the Nélémwa set appears as non-contiguous in Figure 4, in spite of the ideal design of semantic maps in Haspelmath’s terms.

Once it is established – albeit incompletely – a semantic map like Figure 3 constitutes a universal etic grid against which emic categories of specific languages may be described. Each lexical headword (word or root) selects a particular subset out of the total range of potential senses. This is made clear by Figure 4, which converts the data of Table 1 into graphic sets. By analogy with the concept of isoglosses, I propose to call these sets “isolectic sets”.

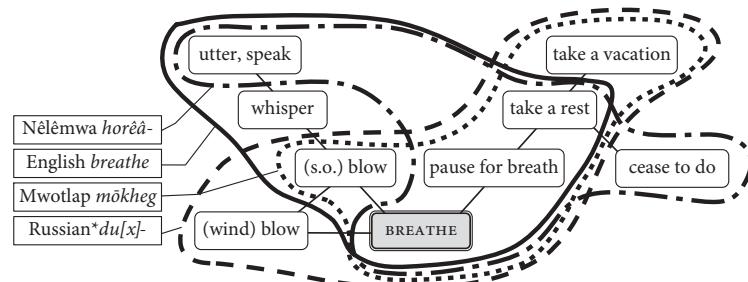


Figure 4. Some isolectic sets around the notion {BREATHE}.

The most instructive point here, in terms of typology, is that the array of cross-linguistic variation, far from being infinite and random, appears to be relatively limited. Of course, the more languages are considered, the more senses will appear in the chart. But even at the small scale of these first observations, the fact that the same patterns of polysemy recur again and again across language families is, in itself, of considerable interest in the search for potential language universals. This sort of cross-linguistic comparison can help see which patterns of polysemy are typologically more common than others (see 3.4.): for example, while the four languages presented here all share the colexification of ⟨breathe⟩ with ⟨take a rest⟩, only one has gone as far as to include the meaning ⟨cease to do⟩. Of course this result with only four languages is not significant; but the possibility of extending the observation to virtually hundreds of languages suggests the sort of research that may be carried out in the future.

### 5.3 Exploring the noun “breath”

The preceding paragraphs have presented the principal cases of colexification associated with the verb “breathe” in my corpus. A much richer semantic network arises if one addresses the domain of nouns. Many languages possess a noun which is cognate with the verb “breathe” (Eng. *breath*) – I will call it here “the {BREATHE} noun”.

#### 5.3.1 “Breath”, “breath of air”, “scent” ...

In some languages, as one would expect, this deverbal noun carries with it part of the polysemy of the verb “breathe”; but most often, languages provide that noun with its own polysemy, which warrants a specific description. Thus, to take the case of Mwotlap, the verb *mōkheg* “breathe; take a rest; be on vacation” has a directly derived noun *nō-mōkheg*, which means equally “breathing”, “rest” and “vacation”. But it also possesses a cognate noun *nō-mōkhe* with its own particular semantics: “breath”, “smell”, “breath of life”, etc.

The literal meaning of the {BREATHE} noun is normally to designate the physical activity, or manner, of breathing (Eng. *pause for breath*; *be short of breath*). In some languages, it also expresses the portion of air inhaled or exhaled during the act of breathing, including its physical properties such as temperature or smell (*hot breath*; *bad breath*). By extension, the same word is sometimes used for all sorts of smells, even when unrelated to an actual process of human breathing: e.g., Mwotlap *nō-mōkhe tētēngé* “the scent (lit. the breath) of flowers”. Through a similar shift between man and nature, the human activity of breathing is sometimes colexified with natural phenomena involving motion of air, such as ⟨breath of air⟩, ⟨wind⟩ or even ⟨cold air⟩.

These different senses seem to be articulated into two chains: on the one hand, a chain ⟨human act of breathing⟩ – ⟨air in motion: breath of air⟩ – ⟨wind⟩ – ⟨cold air⟩; on the other hand, a chain ⟨human act of breathing⟩ – ⟨air coming from human mouth⟩ – ⟨smell coming from human mouth⟩ – ⟨smell, scent in general⟩. To take just one example, Latin *spīritus*, derived from *spīro* “breathe”, is attested with all these meanings (except for ⟨cold air⟩).

#### 5.3.2 “Life”, “spirit”, “mind”, “feelings” ...

But probably the most significant polysemy that is attested with {BREATHE} nouns is the lexical field of “life” and “soul”. This time, among the various properties associated with the act of breathing, the one which is most relevant here is a universal physiological observation: namely, that *the phenomenon of breathing is the most salient property that distinguishes a live creature from a dead body*.

Thus, {BREATHE} nouns or verbs are frequently – perhaps universally – attested in phrases related to the semantic notions of “life” and “death”: see Eng. *breath of life*; *draw one's last breath*; *breathe life into s.th.* ... In Latin, the verb *exspīro* (from *spīro* “breathe”) means literally “breathe s.th. out”, but also serves as a euphemism for “breathe one's last, die” (> Eng. *expire*). Russian *iz-dyxat* “die”, etymologically connected to the root *du[x]*, is exactly parallel to Latin *ex-spīro*.<sup>16</sup>

16. The reason why the sense ⟨die⟩ is not represented on the final semantic map of {BREATHE} (Appendix 2) is because this meaning is always obtained indirectly, through lexical – or

This is how certain languages have come to colexify ⟨breath⟩ and ⟨life⟩. More precisely, the {BREATHE} noun is often related, whether historically or synchronically, with a word whose meaning could be described as “the principle of life, insofar as it can be conceived as specific of an individual”. Indeed, while these languages often possess a separate noun for the abstract concept “life” (Greek *bios*, Lat. *vita*, Arabic *Yāṣa* ...), they also often make use of another term when it comes to embodying this abstract principle, as it were, into an individual being. This is how many – if not all – cultures around the world have elaborated the non-trivial notion of the *soul* or *spirit*: that is, the vital force of an individual, insofar as it is opposed to the inert body.

Needless to say, a wide variety of conceptions can be carried by this notion of *spirit*, depending on cultures, religions, times and people. Despite the risk of simplification, this diversity can perhaps be reduced to a few prototypical concepts. At least, I shall mention here those concepts that are lexified, among the world’s languages, in direct connection with the notion {BREATHE}.

In some languages, the {BREATHE} noun embraces the psychological activity of an individual, in its various manifestations. For example, Classical Latin *animus*<sup>17</sup> is attested with the following meanings: ⟨vital principle of an individual: soul⟩; ⟨seat of reason and intelligence: mind⟩; ⟨seat of will and desire: will⟩; ⟨seat of feelings and passions: “heart”⟩; ⟨seat of courage and vital energy⟩; ⟨strong passions: pride⟩...

The semantic range is not necessarily as wide as this, and is sometimes restricted to just a certain type of feeling. To take another Latin example, the noun *spiritus*, besides its other meanings mentioned in 5.3.1., is also attested with psychological senses; but as far as Classical Latin is concerned, these are essentially restricted to ⟨pride, arrogance, self-importance⟩. During the later history of Latin and of Romance languages, the set of psychological meanings related to *spiritus* has enriched considerably. Thus, French *esprit* has a wide polysemy of its own, which includes ⟨mind, thought⟩, ⟨intelligence⟩, ⟨wit⟩, ⟨seat of feelings⟩, ⟨character, moral disposition⟩, ⟨frame of mind, mood⟩. A few phrases illustrate these senses, such as *garder à l'esprit* “keep in mind”, *avoir l'esprit vif* “have a quick mind”, *avoir de l'esprit* “to be witty”, *avoir l'esprit à rire* “to be in

---

phraseological – derivation, but never directly (“strict colexification”). For obvious reasons, no language is found where ⟨die⟩ and ⟨breathe⟩ are expressed by exactly the same form in synchrony. As a principle, those senses which are attested nowhere in strict colexification with the pivot notion do not qualify for inclusion in its semantic map (see 3.3.).

17. Admittedly, *animus* did not have ⟨breathe⟩ nor ⟨breath⟩ among its senses in the “synchrony” of Classical Latin. However, it is etymologically linked to Greek *anemos* “wind” and Sanskrit *aniti* “breathes”; and more importantly, it is closely cognate with the noun *anima*, whose wide polysemy does include ⟨breath⟩ and ⟨wind⟩. As a result, I take *anima* as the relevant headword for Latin (see 7.1.3.); *animus* is only included in the corpus by virtue of its synchronic cognacy with *anima* (“loose colexification”).

a mood for laughing”, *dans l'esprit de l'époque* “in the spirit of the age”, *esprit d'équipe* “team spirit”, *retrouver ses esprits* “to collect one’s wits” ... Incidentally, because French *esprit* – like Eng. *spirit* – no longer shows any connection with {BREATHE} in synchrony, it can only be included in our corpus on a historical basis. In case we want to restrict our observations to synchronical polysemies, then the examination of this root should be restricted to Classical Latin *spiritus*, whose semantic array is already wide (7.1.3.).

Similar semantic extensions can be found in other languages, including in the form of synchronically coexisting senses. For example, it is remarkable that Standard Arabic also translates some of the psychological senses of Fr. *esprit* with *rūḥ* a noun related to {BREATHE} (7.1.10.): e.g., *rūḥ al-taṣāun* “team spirit”, *al-rūḥ al-ḥarbiya* “warlike spirit”. The polysemy of Russian *dux* also presents similar characteristics in synchrony – even more if one considers the whole set of words that form the cognate set of the root \**du[x]*<sup>18</sup> (7.1.4.).

### 5.3.3 “Soul”, “spirit”, “supernatural being” ...

The group of senses just reviewed (⟨spirit⟩, ⟨mind⟩, ⟨character⟩ ...) forms a branch of its own in the semantic map of {BREATHE}, covering the domain of psychological and mental qualities of the socialized person. It should be carefully distinguished from another concept: the soul. The semantic nuance is familiar to all Latinists, since it is formally distinguished in Latin as (masculine) *animus* vs. (feminine) *anima*. While *animus* describes the various faculties, feelings and emotions of individuals in their social activities, *anima* has a deeper existential meaning, as it refers to the primal faculty of being alive – see also the derived noun *animal* “living being”. Therefore, nouns like *anima* will be typically used in contexts dealing not with social behaviour, but with death. In this perspective, the soul can be described as “that part of an individual which leaves the body when death comes.” Depending on the cultural context, this separation from the body will be understood either as the complete disappearance of the soul, or, on the contrary, as its survival in different forms: migration of the soul to an invisible abode of the dead, restless wandering as a ghost in the present world, reincarnation (metempsychosis) into a new human body, or metamorphosis into a supernatural being.

One may think that these cultural issues are not relevant for our linguistic study, but they are. Only the understanding of such religious beliefs makes it possible to define a satisfactory semantic path between, on one end of the semantic chain, the notion of breathing, and on the other end, the representation of ghosts and other supernatural

---

18. Amongst the various lexical items that are etymologically related with this root, the noun *duša* “soul, spirit ...” has received special attention in Wierzbicka (1992: 31ff.).

beings, whether in an animist or a monotheist context. This polysemy can be illustrated again with Fr. *esprit* (<*spíritus*): besides the mental and moral senses used in a social context (taking over the semantics of *animus*), it can also refer to the soul of a living being (*anima*), including in the form of a ghost (e.g., *croire aux esprits* “believe in ghosts”). Finally, *esprit* can designate any supernatural being of divine nature (*l'esprit du fleuve* “the spirit of the river”), whether good (*esprits célestes* “heavenly spirits”) or evil (*esprit malin* “evil spirit”); and in the context of a monotheist religion, the same word may even come close to referring to the supreme divinity *par excellence*, as in *le Saint-Esprit* “the Holy Spirit”.

This impressive range of “spiritual” meanings is not exclusive to the lexicon of Latin (*animus, anima, spíritus*) and of its daughter languages. Surprisingly similar patterns of polysemy are found elsewhere: Greek *psúkhē* and *pneuma*; Sanskrit *ātman*; Russian *du[x]*; Arabic *rūh* and *nafs*; Aleut *anri*; Nahuatl *imi’yo*; and so forth (see Appendix 1).

#### 5.3.4 Going from “breath” to “supreme spirit”

To be precise, there are two ways one could account for the inclusion of supernatural beings in the semantic map of {BREATHE}. One hypothesis would involve a generalization process, whereby the soul of a human individual, insofar as it is said to survive after death in the supernatural form of a ghost, would serve as a model for all other supernatural creatures, even when they do not originate in a deceased person. In this case, the likeliest semantic chain would be:

⟨breath⟩ – ⟨⟨breath of life⟩ – ⟨vital force of an individual, s.o.’s spirit⟩ – ⟨immortal part of an individual that survives death: soul⟩ – ⟨s.o.’s ghost⟩ – ⟨supernatural being, even when not of human origin; a spirit, good or evil⟩

The likeliness of this scenario is confirmed by the existence of similar semantic shifts with other lexemes, though unrelated to {BREATHE}. For example, the Mwotlap noun *na-tmat* (François, in PREP.), etymologically “dead person”, is a polysemous word that colexifies ⟨deceased person⟩ – ⟨wandering soul of a deceased person, ghost⟩ – ⟨monster, spirit; any supernatural being, whether good or bad⟩ – ⟨the Biblical Devil⟩.

A second hypothesis would make a shortcut between the very act of breathing and the notion of divinity, with no need to posit ⟨soul, ghost⟩ as a missing link. Indeed, in many cultures, the immortality of divine entities is metaphorically compared with an invisible breath of air, a magic wind. This divine wind may sometimes be “blown into” a thing or a person to endow it with holiness or supernatural power. This metaphor, for example, underlies the use of Eng. *inspiration* (for an artist, a poet, a prophet) from Latin *inspiro* “blow into”. Likewise, the Classical Greek *pneuma*, literally “breath, breath

of air ...” is attested with the meaning “divine breath”<sup>19</sup> but never with the sense “soul” or “ghost”. Finally, a process of metonymy triggers the shift from ⟨divine breath⟩ to ⟨the divine entity or supernatural being from whom a divine breath emanates⟩. These examples would therefore rather advocate for a second semantic path:

⟨breath of air emanating from a human person⟩ – ⟨divine breath: supernatural power emanating from an immaterial entity⟩ – ⟨supernatural being exhaling divine breath, divine spirit⟩

Because both chains seem to be semantically likely and empirically grounded, I prefer not to choose between them (see discussion in 4.2.). Such ambiguity is not necessarily an issue, and may well depend on the specifics of each language or culture. It can be easily represented on the typological map of {BREATHE} by drawing two distinct paths leading from ⟨breathing⟩ to ⟨supernatural being⟩: see Figure 5 in Appendix 2.

#### 5.3.5 From “soul” to reflexive marking

Finally, a further extension from the sense ⟨soul⟩, ⟨spirit⟩ is the designation of an individual’s “person”, “essence” or “ego” – what one may define as one’s inner, deeper identity, as opposed, for example, to one’s social representation. This more or less corresponds to the semantics of English *self*.

Even more interestingly, this quite abstract meaning has sometimes grammaticalized into a reflexive marker, in a way precisely parallel to English ⟨know⟩ *your self* > ⟨know⟩ *yourself*. This semantic path is witnessed in three languages in my corpus. In Sanskrit (7.1.1.), the famous concept *ātman* (etymologically “breathing”, from *an-* “breathe”) has a wide semantic array, going from ⟨breath of life⟩ to ⟨vital force⟩, ⟨soul⟩ and ⟨the self, the abstract person⟩ as well as ⟨essence, peculiarity (of something)⟩. But one of its principal uses in texts seems to be as a grammatical marker for reflexive; this is especially clear from the list of dozens of compounds based on *ātma-* (of which only a short selection is given in the Appendix), e.g., *ātma-jñā* “knowing one’s self”, *ātmēśvara* “master of one’s self”, *ātma-ghāta* “suicide”, *ātma-grāhīn* “selfish”...

Likewise, the ordinary reflexive marker for Standard Arabic is *nafs-i* (1SG possessed form of *nafs*, parallel to Eng. “my-self”). This is in fact a noun *nafs* meaning ⟨soul⟩,

19. Historically speaking, this specific sense, despite being already attested in Plato’s works, was later spread by the Septuagint in their translation of the Bible. Whereas the noun *psúkhē* “soul, spirit” had lost its etymological relationship to “breath”, the noun *pneuma* was still synchronically the word for “breath, blow of air”: this is probably why it was chosen to translate Hebrew *ruach* “breath, air; strength; wind; spirit; courage; temper; Spirit” (Vine 1985: 240; see the cognate Arabic *rūh* in the appendix). Exactly in the same way, the semantic calque took place in Latin with *spiritus* “breath, blow of air; soul ...” rather than *animus*, because the connection of the latter noun with “wind, breath” was then no longer perceptible.

*(essence, being, abstract person), (self), (mind, psyche) as well as (the same) ... And crucially, this whole semantic array is closely connected – via loose colexification – with the noun *nafas* “breathing, breath, breath of life” (root *n.f.s*). This example confirms the relevance of a semantic chain (breathing) – (breath of life) – (vital force) – (person, self) – (reflexive).*

The other Arabic root with a similar polysemy, *r.w.h* apparently does not go that far, at least for Standard Arabic. However, Naim (2007: 315) reports the grammaticalization of *rūh* as a reflexive marker in modern Yemeni Arabic (as in *ʔalattim rūh-i* “I’m hitting myself”).<sup>20</sup> This confirms the potential bridge between lexicon and grammar, which is potentially present within this lexical field of {BREATHE}.

#### 5.4 Drawing a more complete map for “breathe”

The previous sections have surveyed the main patterns of polysemy, regarding both the verb “breathe” and the cognate noun “breath”, based on a corpus of 16 lexical headwords (covering 114 words altogether) in 13 languages. These observations result first in a comparative sense chart (Table 2 in Appendix 2), and in the typological map of {BREATHE} (Figure 5).<sup>21</sup>

The reader will find in Appendix 3 a representation of eleven significant lexical headwords of the corpus, in the form of “isolectic sets” (see 5.2.). Crucially, these figures show clearly how the universal semantic map was carefully drawn on an empirical basis. Indeed, following the methodological principles stated in 4.2., senses must be organized on the universal map so that each isolectic set covers a *contiguous* part of the map; and every semantic chain proposed, based on functional or ontological properties, must be confirmed empirically by the existence of such polysemous chains in actual languages. With just one exception already noted (fn.16), these two methodological requirements are rigorously fulfilled by the universal map I propose.

The interest of this typological map lies both in its complexity and its simplicity. First, knowing how universal the activity of breathing is, it is impressive to see how each language has proven capable of evolving its own way, bringing about highly sophisticated, culture-specific vocabulary such as “the self”, “divine inspiration”, or “be on vacation”. Yet, by the same token, probably even more instructive is the extent to which remote languages can follow just the same semantic paths, well beyond genetic

20. Technically, although this is an instance of the noun *rūh*, this specific polysemy should count as loose rather than strict colexification (see 3.3.) in the map of *rūh*, because it involves the same lexeme across two distinct *états de langue*: the chain (breath) – (soul) – (spirit) – (person) ... belongs to Standard Arabic, but the grammaticalization (person) – (reflexive) to Yemeni Arabic. See Figure 15.

boundaries and historical times. Thanks to this type of typological survey, certain metaphors sometimes believed to be specific of certain civilizations (e.g., the connection “breath” – “soul” – “spirit” found in the Bible) can appear to be in fact widespread among the world’s cultures. It is probable that lexical typology, as much as grammatical typology, will tell us a lot about the universality of our perceptions and feelings, and about the unity of mankind.

#### 6. Conclusion

Of course, such a semantic map is by no way comprehensive; it will always be possible to improve it by increasing the number of languages considered. Rather, the objective of the present overview was essentially to explain and illustrate a possible method for undertaking research in lexical typology. My objective was to find a satisfying balance between the two conflicting demands at stake in language typology: the search for universals vs. the respect for each language’s uniqueness. On the one hand, typological comparison requires that linguistic data be observed from a universalist angle, through the definition of language-independent, functionally-based criteria that could be observed – or at least looked for – in potentially any language of the world. On the other hand, the need for comparison should not sacrifice the subtle nuances that make each language unique. Hopefully, the Colexification Model proposed here, based on the definition of minimal semantic atoms and the observation of their interactions in the various languages of the world, should provide a satisfying balance between language-specific analyses and a more universal approach.

More issues still deserve to be addressed. For example, the question of diachrony, and specifically of the directionality of semantic change, could be researched in the future. Also, the representation of the data could be improved, e.g., by using three dimensions instead of two, or by adding various attributes for each semantic connection – distinguishing metaphors from metonymies, or statistically frequent cases from rarities... Finally, little has been said here about the possible applications of this model, whether in a universalist, cognitivist perspective, or in the reconstruction of historical change in particular language groups.

Obviously, the domain of lexical typology still provides ground for substantial debate and reflection, from both a theoretical and a practical standpoint. But while we pursue these necessary discussions, we must not forget to collect the raw material for this research – namely, fine-grained lexical data gathered from various parts of the world. This work of documentation and analysis is especially urgent for endangered languages, most of which have so far received too little attention from lexicographers.

## 7. Appendices

### 7.1 Appendix 1: Lexical data

#### 7.1.1 Sanskrit

Source: Monier-Williams 1970 [1899]; Stchoupak et al. 1987 [1932].

- Lexical item: *ātman*.
  - 1. breath. 2. (breath of) life; principle of life, vital force. 3. the individual soul, spiritual force of the person. 4. the self, abstract individual; oneself (reflexive pronoun), one's own. 5. the person, esp. body. 6. understanding, intellect, mind. 7. essence, character, peculiarity. 8. effort, firmness. 9. highest personal principle of life, Brahma.

Cognate, derived and compound forms (selection):

<i>ātma-vat</i>	animated, having a soul.	<i>ātma-bhāva</i>	1. existence of the soul. 2. the self, proper or peculiar nature.
<i>ātma-</i>	1. soul. 2. self, one's own...	<i>ātma-yoga</i>	union with the supreme spirit.
<i>ātma-grāhin</i>	taking for one's self, selfish.	<i>ātma-vīd</i>	knowing the nature of the soul or supreme spirit.
<i>ātma-ghāta</i>	suicide.	<i>ātma-sáni</i>	granting the breath of life.
<i>ātma-jñā</i>	1. knowing one's self. 2. knowing the supreme spirit.	<i>ātma-dhīna</i>	1. depending on one's own will. 2. one whose existence depends on the breath or on the principle of animal life: sentient.
<i>ātma-jyotis</i>	the light of the soul or supreme spirit.		
<i>ātma-tā</i>	essence, nature.		
<i>ātma-dā</i>	granting breath or life.		
<i>ātma-pāta</i>	descent of the soul, re-birth.	<i>ātmē-śvara</i>	master of one's self.

Cognate form: *an* (3sg á̄n-iti)

(cognate with Greek *anemos* “wind”, Latin *anima*)

- 1. breathe, respire. 2. gasp. 3. live, be alive. 4. move, go (?).

Cognate, derived and compound forms (selection):

<i>aná</i>	breath, respiration.	<i>ánila</i>	1. air, wind. 2. the god of wind.
<i>aná-vat-va</i>	the state of being endowed with breath or life.		3. wind as one of the humors of the body. 4. rheumatism.
<i>anana</i>	breathing, living.		

#### 7.1.2 Classical Greek

Source: Bailly (1950 [1894]).

- Lexical item: *psūkhē*.
  - 1. (s.o.'s) breath, puff of air. 2. breath of life, vital force. 3. (s.o.'s) life. 4. living being; person. 5. darling. 6. soul (vs. body): seat of feelings and passions, heart. 7.

(s.o.'s) moral disposition, character; nature (of s.th.). 8. seat of intelligence, mind. 9. seat of will and desire. 10. soul separated from body and surviving in hell; ghost. 11. butterfly.

Cognate, derived and compound forms (selection):

<i>psūkhikos</i>	1. vital. 2. living being; animal. 3. terrestrial, material. 4. of the soul, spiritual.	<i>psūkhos</i>	3. cool down (s.th.). 4. get cold; fall, die.
<i>psūkhō</i>	1. breathe, blow air. 2. breathe out, reject.	<i>psūkhros</i>	1. fresh breath of air; cold air, coldness. 2. winter. 1. cold. 2. sterile. 3. vain, useless. 4. lifeless. 5. indifferent, impervious.

- Lexical item: *pneuma*.

1. breath of air; wind. 2. act of breathing; (s.o.'s) breath. 3. sound (of flutes). 4. breath of life, life. 5. breathlessness. 6. smell, scent. 7. enthusiasm, energy, fervor; wrath. 8. divine breath, divine power. 9. spirit, supernatural being, whether good (angel) or bad (devil); Holy Spirit. 10. aspiration (phonetics).

Cognate form: *pneō*.

1. (wind) blow. 2. (s.o.) breathe, blow. 3. be alive. 4. be in a particular moral disposition (pride, anger, arrogance ...). 5. exhale a smell, smell (good or bad). 6. play the flute. 7. (*passive*) be inspired, be wise.

#### 7.1.3 Classical Latin

Source: Gaffiot (1934).

- Lexical item: *anima*.

(cognate with Greek *anemos* “wind” Skr. *aniti* “breathes”)

1. air in motion, breath of air. 2. act or manner of breathing. 3. breath (good or bad). 4. (breath of) life; principle of life, vital force. 5. being, creature, person. 6. darling. 7. soul (opp. body) that survives death; souls of the dead.

Cognate form: *animus*.

1. principle of life (opp. body). 2. mind, thought, seat of intelligence. 3. opinion, thought. 4. seat of will and desire; will, intention. 5. seat of feelings: soul, heart. 6. feelings, emotions, passions. 7. frame of mind, mood. 8. courage, energy, fervor, pride, arrogance. 9. darling.

Cognate, derived and compound forms (selection):

<i>animal</i>	living being, animal.
<i>animōsus</i>	1. courageous, bold. 2. proud. 3. ardent.

- Lexical item: *spīritus*
  - air in motion, breath of air.
  - act or manner of breathing; breath.
  - sigh.
  - smell, scent.
  - (breath of) life; principle of life.
  - divine breath; magic or poetic inspiration.
  - soul (opp. body).
  - self, person.
  - pride, arrogance, self-importance.
  - feelings, state of mind, moral disposition.

Cognate form: *spīro*.

- (wind) blow.
- (sea) bubble.
- breathe.
- be alive.
- be inspired.
- (s.o., s.th.) exhale a smell.
- breathe out (s.th.).
- (fig) exude [cruelty ...].

#### 7.1.4 Russian

Sources: Sakhno (2005: 89 ff.); Pauliat (1991).

- Lexical item: *dux*.
  - breathing, breath.
  - breath of life.
  - spirit (of s.o./s.th.): moral disposition, frame of mind.
  - mood (good or bad).
  - morale, courage.
  - supernatural being (good or evil); God (*Syjatoj dux* "Holy Spirit").
  - ghost.

Cognate, derived and compound forms (selection):

<i>dut'</i>	1. (s.o./wind) blow. 2. drink a lot.	<i>otdyx</i>	1. pause, rest. 2. leisure, vacation.
<i>zaduvat'</i>	1. (wind) start blowing.		
	2. (s.o.) blow (candle+).	<i>peredyška</i>	short pause, respite.
<i>vozdux</i>	1. air. 2. open space, outside.	<i>dušok</i>	bad smell.
<i>dysat'</i>	breathe.	<i>duxi</i>	perfume.
<i>doxnut'</i>	breathe, blow.	<i>duša</i>	1. soul, spirit. 2. seat of feelings, heart. 3. inhabitant, person.
<i>dyxanie</i>	breathing, breath.		
<i>dyxatel'nyj</i>	respiratory.		
<i>dušit'</i>	strangle, suffocate; oppress.	<i>duševnyj</i>	1. psychic, mental.
<i>uduše</i>	breathlessness, asthma.		2. sincere, cordial.
<i>doxnut'</i>	die.	<i>duxovnyj</i>	spiritual; holy, sacred; ecclesiastical.
<i>izdyxat'</i>	die.		
<i>vzdyxat'</i>	sigh.	<i>duxovenstvo</i>	clergy.
<i>otdyxat'</i>	take rest.	<i>vdoxnenie</i>	(poetic/magic) inspiration, enthusiasm.
		<i>oduševlennyj</i>	animate.

#### 7.1.5 Mandarin Chinese

Source: [no author] (1990); [no author] (1996).

- Lexical item: *qi*.
  - weather, atmosphere.
  - gas.
  - air.
  - (s.o.'s) breath.
  - smell (good or bad), scent.
  - (s.o.'s) manner, ways, attitude, style.
  - (s.o.'s) spirits, moral strength, morale.
  - energy, vital force, vital breath.
  - annoy, irritate (s.o.).

Cognate, derived and compound forms (selection):

<i>qihòu</i>	weather, climate.	<i>qiwèi</i>	1. smell, scent. 2. (fig) taste, style, fashion.
<i>qixiāng</i>	1. meteorology. 2. (fig) atmosphere.	<i>qipài</i>	style; stylishness.
<i>qichuān</i>	breathless, asthmatic.	<i>qipò</i>	1. character strength; boldness, daring; will. 2. majesty.
<i>qixī</i>	1. breath; last breath. 2. smell, scent. 3. (fig) taste, style, fashion.	<i>qiyàn</i>	arrogance, insolence.
<i>qishi</i>	1. strength, vigor, energy.	<i>qigài</i>	manner, ways.
	2. momentum, impetus.	<i>qizhì</i>	character, (good) moral qualities.
<i>qili</i>	strength, vigor, energy; effort.	<i>qijié</i>	honesty, frankness.
<i>qināo</i>	get angry.	<i>qiliàng</i>	open-mindedness, tolerance.
<i>qifēn</i>	furious, angry, exasperated.	<i>qinèi</i>	disheartened, depressed.
<i>qihuà</i>	angry words; words ( <i>huà</i> ) uttered in a fit of anger.	<i>qxīng</i>	character, disposition; mood.
		<i>qishèng</i>	be in a bad mood.
<i>qishi xiōngxiōng</i>	fierce, furious, arrogant.		

#### 7.1.6 Inuit/Aleut

Eskimo-Aleut family. Spoken in Greenland and Alaska.

Source: N. Tersis (pers. comm.); Fortescue et al. (1994).

- Lexical item: (Inuit) *ani-* "breathe, blow".

Cognate, derived and compound forms:

<i>aniqniq</i>	breathing; breath.	<i>aniqnii-q-pu-q</i>	breathe one's last, die.
<i>aniqsaaqtuq-puq</i>	breathe.	<i>aniqsaaq</i>	spirit, ghost.

Cognate form: (Aleut) *anr(i)* "breathe, blow".

- breath.
- voice.
- principle of life, life.
- spirit, soul.
- ghost.

Cognate form: *anrari* "be alive".

#### 7.1.7 Nahuatl

Spoken in Mexico. Source: Marie-Noëlle Chamoux (pers. comm.).

- Lexical item: *imi'iyo ~ i'iyac*.
  - breath.
  - smell (esp. bodily smell).
  - V + imi'iyo* "take rest".

#### 7.1.8 Mwotlap

Austronesian; Oceanic subgroup. Spoken in Vanuatu. Source: François (in PREP.).

- Lexical item: *nō-mökhe*.
  - (s.o.'s) breath.
  - breath of life, life; principle of life.
  - smell (good or bad), scent (of s.o./s.th.).

Cognate form: *mōkheg*.

1. breathe; breathe into. 2. perceive a smell. 3. pause, take rest; be on vacation; be retired.

#### 7.1.9 *Nélémwa*

Austronesian; Oceanic subgroup. Spoken in New Caledonia. Source: Bril (2005)

- Lexical item: *horêā-t*.
  1. (s.o.'s) breath, breathing. 2. breath of life, life. 3. spoken message.

Cognate form: *horêān*.

1. breathe. 2. be alive. 3. pause, take rest. 4. cease to do.

#### 7.1.10 *Standard Arabic*

Source: Reig (1983); Naïm (2007).

- Lexical item: *rūh* [root *r.w.h*]
  1. breath of life. 2. soul, soul of the dead; mind; spirit. 3. supernatural power, spirit (good or evil); divinity. 4. character, moral disposition; spirit. 5. morale, mental strength. 6. perfume essence, alcohol.

Cognate, derived and compound forms (selection):

<i>rūḥī</i>	1. spiritual. 2. alcoholic.	<i>istirāḥa</i>	take a rest, relax, be quiet.
<i>rūḥānī</i>	spiritual; divine; immaterial; sacred.	<i>rāḥa</i>	rest, quietness; ease, comfort.
<i>arwāḥiyah</i>	animism.	<i>rūḥ-i</i>	lit. “my soul” > grammaticalized as a reflexive marker (“myself”) in modern Yemeni Arabic
<i>rīḥ</i>	breath of air, wind.		
<i>rāʔiḥah</i>	smell, scent (good or bad).		
<i>mirwāḥah</i>	fan, propeller.		
<i>rawwāḥa</i>	ventilate, air; put scent in.		

- Lexical item: *nafas* [root *n.f.s*]
  1. breathing. 2. puff of air, breath. 3. sip. 4. breath of life.

Cognate form: *nafs*.

1. soul, vital force of the individual. 2. essence, being, the person itself. 3. the self; Reflexive marker (myself, yourself ...). 4. the same. 5. psyche; psycho-. 6. jinx, curse on s.o.

Cognate, derived and compound forms (selection):

<i>tanaffasa</i>	1. breathe, blow. 2. be reassured.	<i>nafsiya</i>	psychology.
<i>tanaffus</i>	breathing, breath.	<i>naffasa</i>	comfort, appease.
<i>nafṣī</i>	psychic, psychological.	<i>tanāfasa</i>	compete, rival

#### 7.1.11 *Beja*

Afro-asiatic family. Spoken in Sudan.

Source: Martine Vanhove (pers. comm.), after Roper (1928).

- Lexical item: *šūk*.
  1. breathing, breath. 2. soul.

#### 7.1.12 *Makonde*

Niger-Congo; Bantu subgroup. Spoken in Tanzania. Source: Sophie Manus (PERS. COMM.)

- Lexical item: *ku-pumula*.
  1. breathe. 2. take rest.

#### 7.1.13 *Sar*

Nilo-Saharan; Sara–Bongo–Baguirmian subgroup. Spoken in Chad and Sudan.

Source: Pascal Boyeldieu (PERS. COMM.)

- Lexical item: *koo*.
  1. breathing, breath. 2. air, gas.

taa koo/take/breath/“pause for breath, take a rest”

## 7.2 Appendix 2: Results and semantic map

### 7.3 Appendix 3: Some isolectic sets for “breathe”

The following pages represent eleven lexical entries (out of the sixteen of the corpus) in the form of “isolectic sets” (see 5.2.). The universal map presented in *Appendix 2* is reproduced identically for each language, and used as a visual etic grid against which the emic categorizations made by each language are visualized.

Each isolectic set consists of two levels. The greyed area with a solid line represents “strict” colexification: it shows the semantic contour of the lexical entry itself (the one in the title). On the other hand, the dotted line allows supplementing this first area with indirect or “loose” colexification – generally, other forms in the same language cognate with the lexical entry (see 3.3.). These isolectic sets are further commented upon in 5.4.

Table 2. Lexical data on the polysemy of {BREATHE}

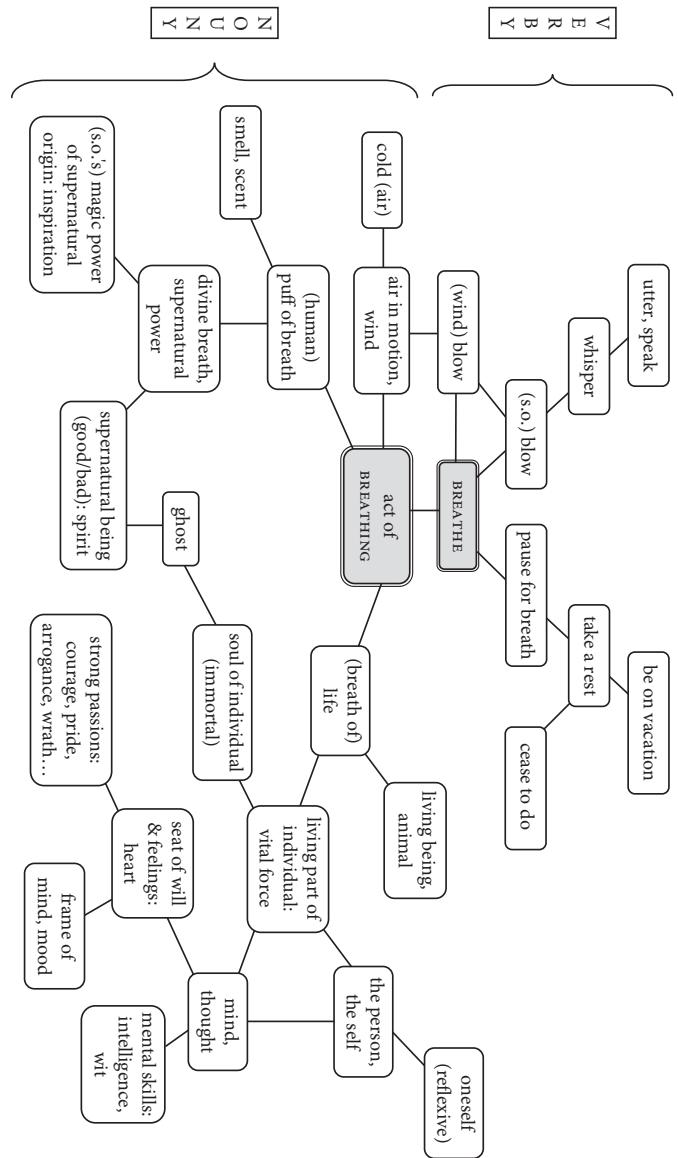
V  
E  
R  
YN  
O  
U  
Y

Figure 5. The semantic map of {BREATHE}.

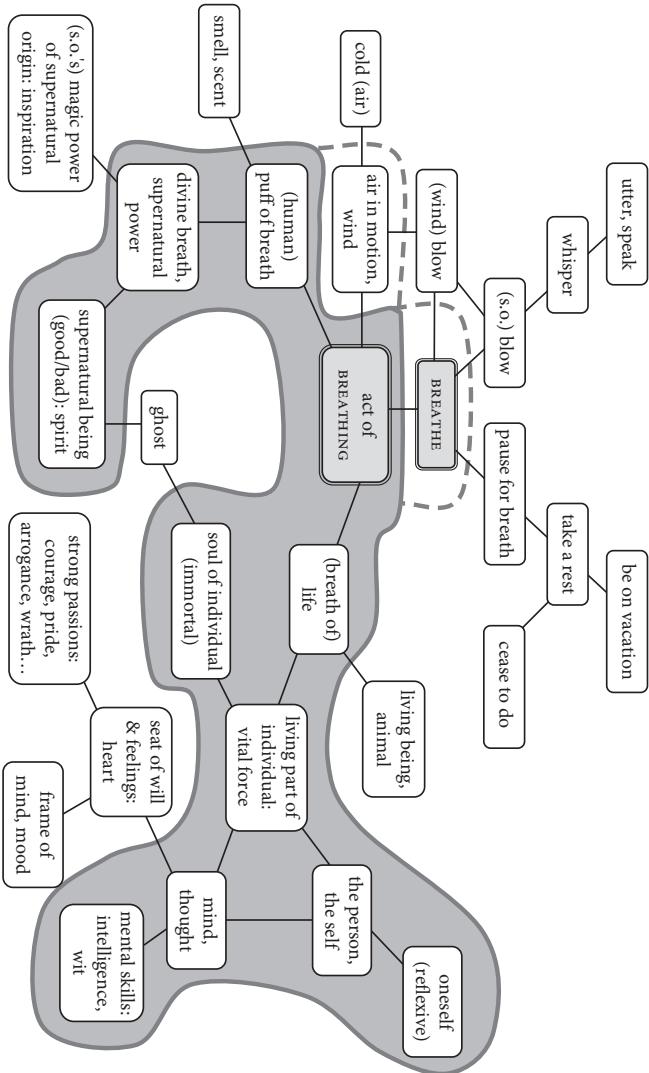
Figure 6. Isolectic set for Sanskrit *ātman*.

Figure 7. Isolectic set for Greek *psūkhē*.

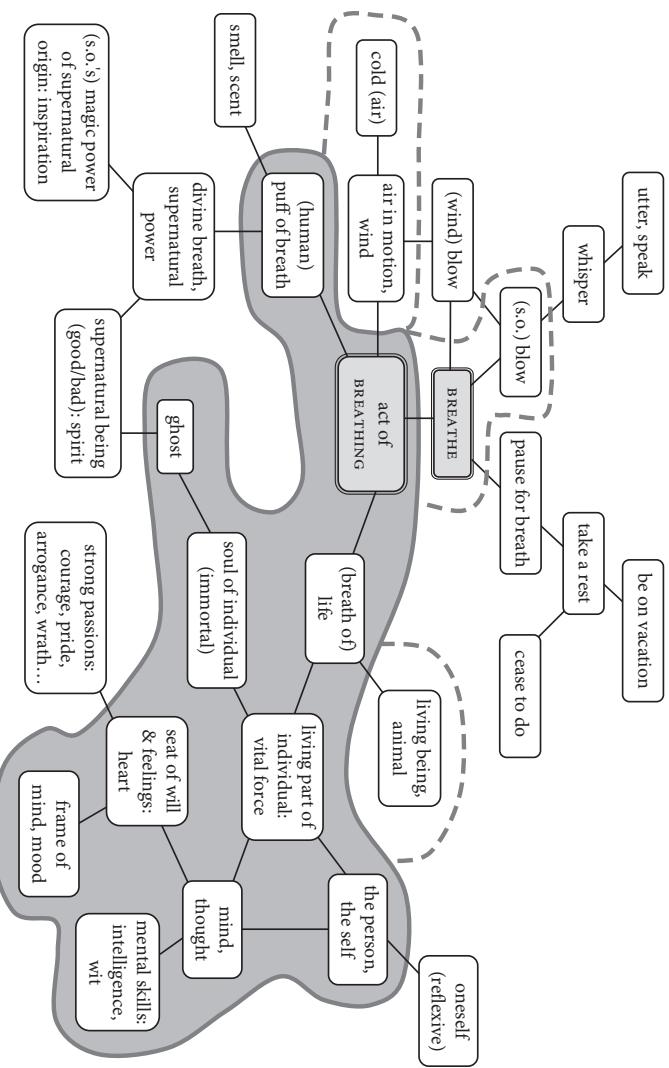
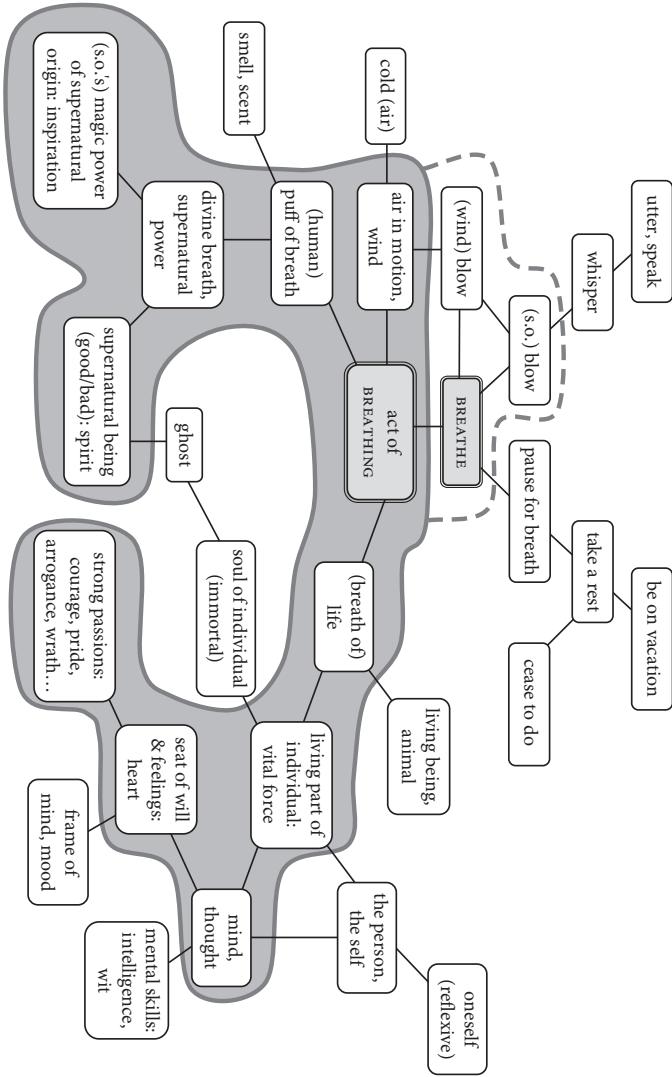


Figure 8. Isolectic set for Greek *pneuma*.



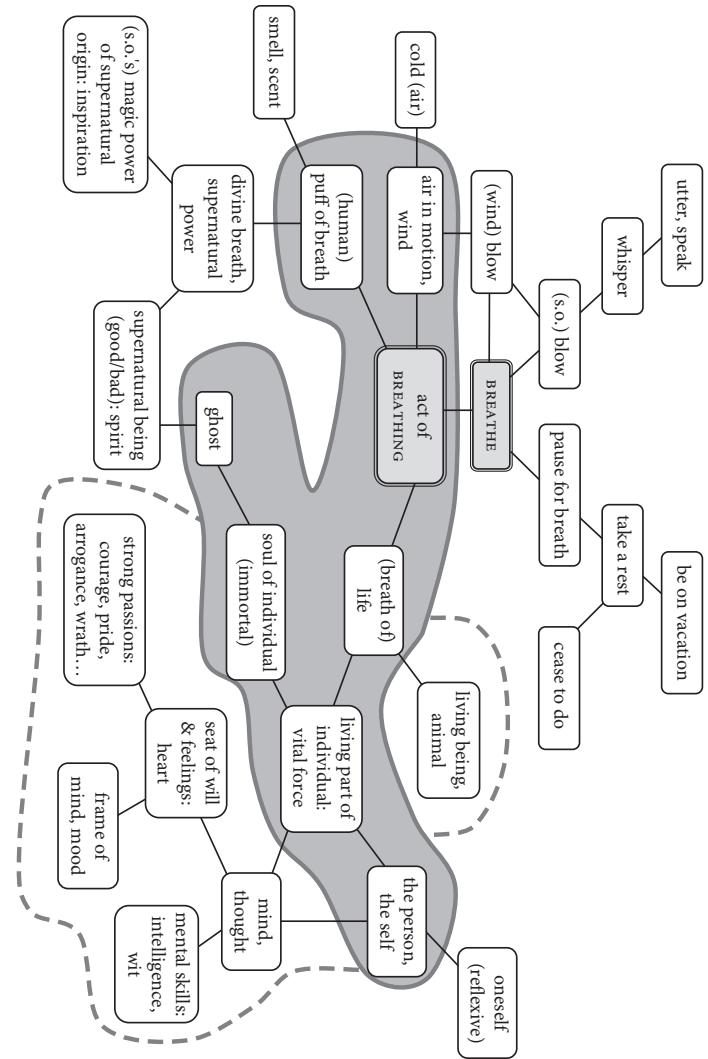
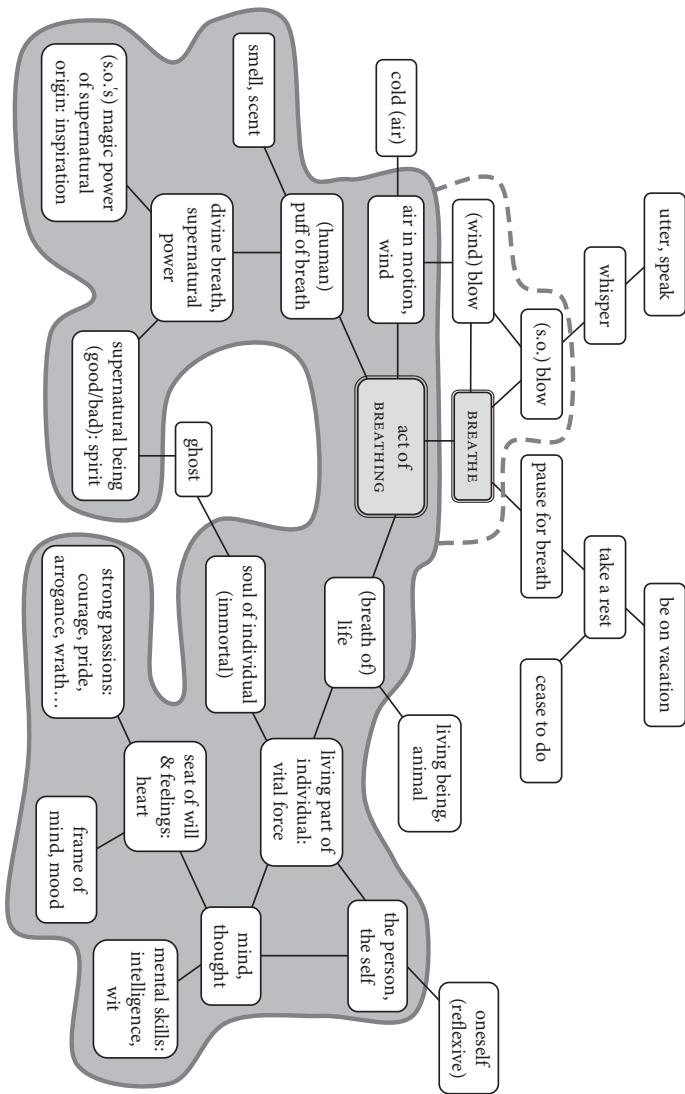
Figure 9. Isolectic set for Latin *anima*.Figure 10. Isolectic set for Latin *spiritus*.

Figure 11. Isogloss set for Russian *dux*.

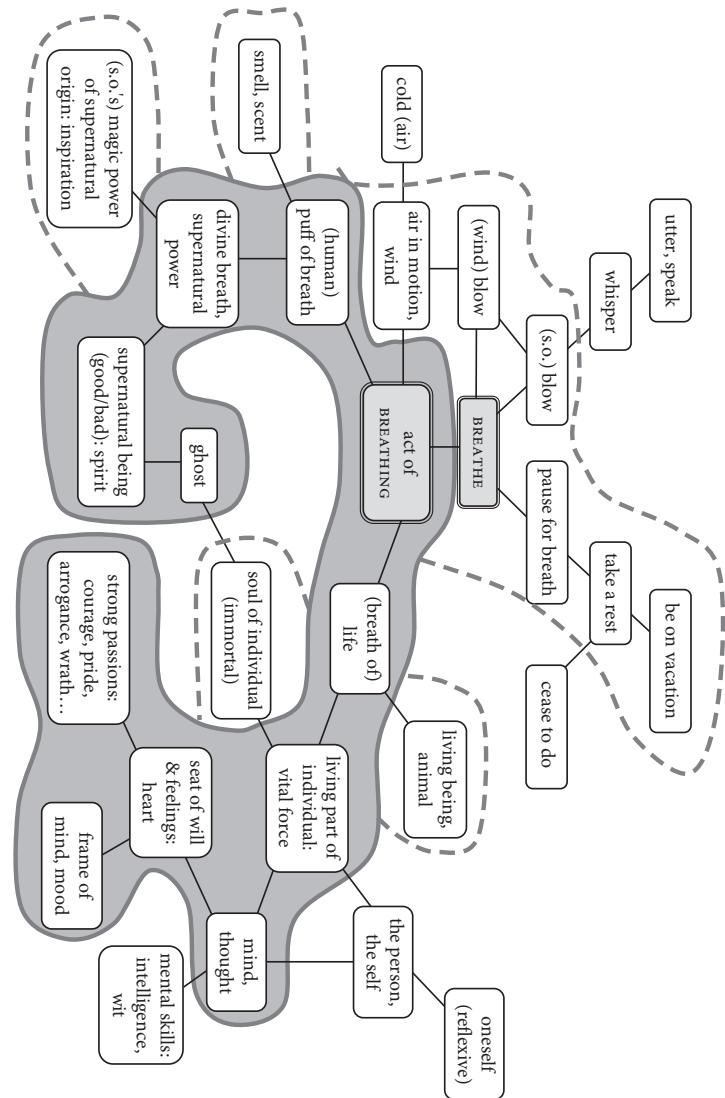
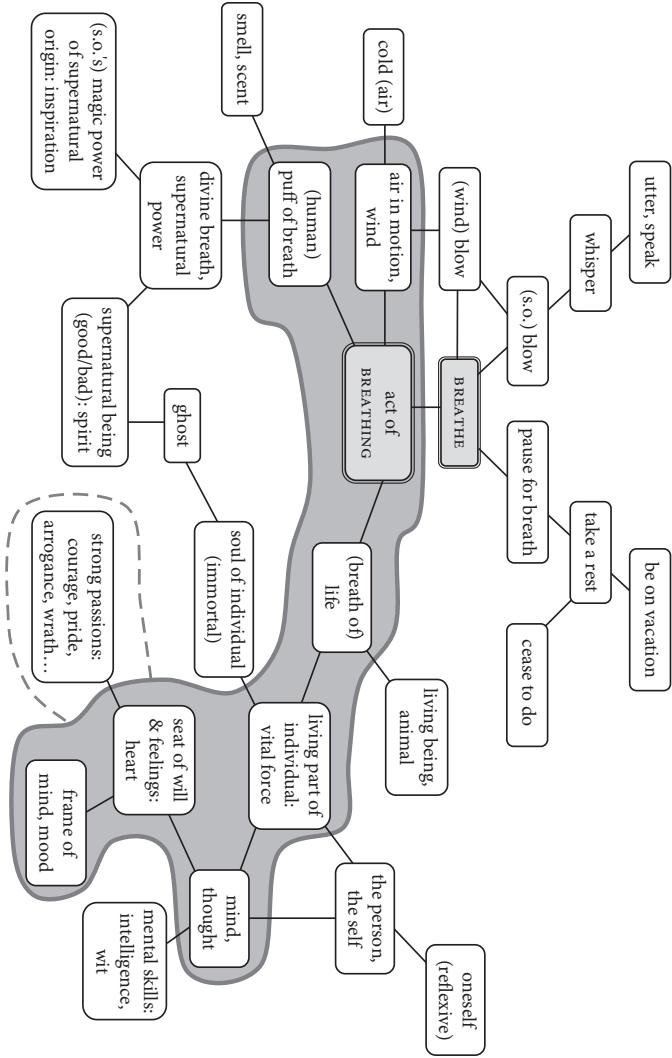


Figure 12. Isolectic set for Chinese *qi*.



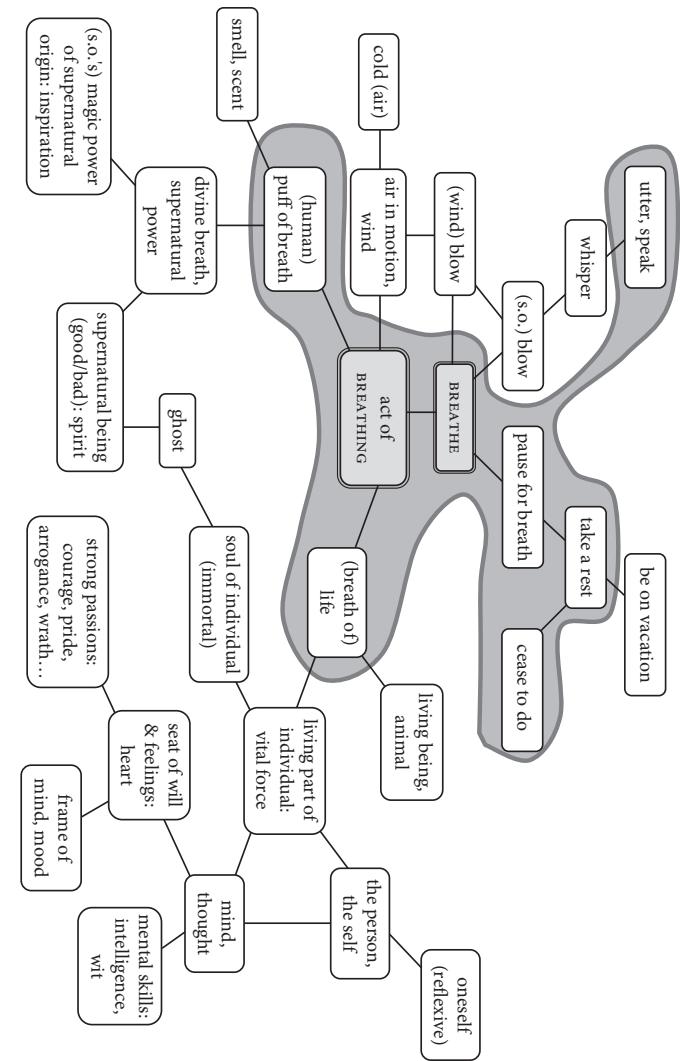


Figure 13. Isolectic set for Nálenwa hórelə-.

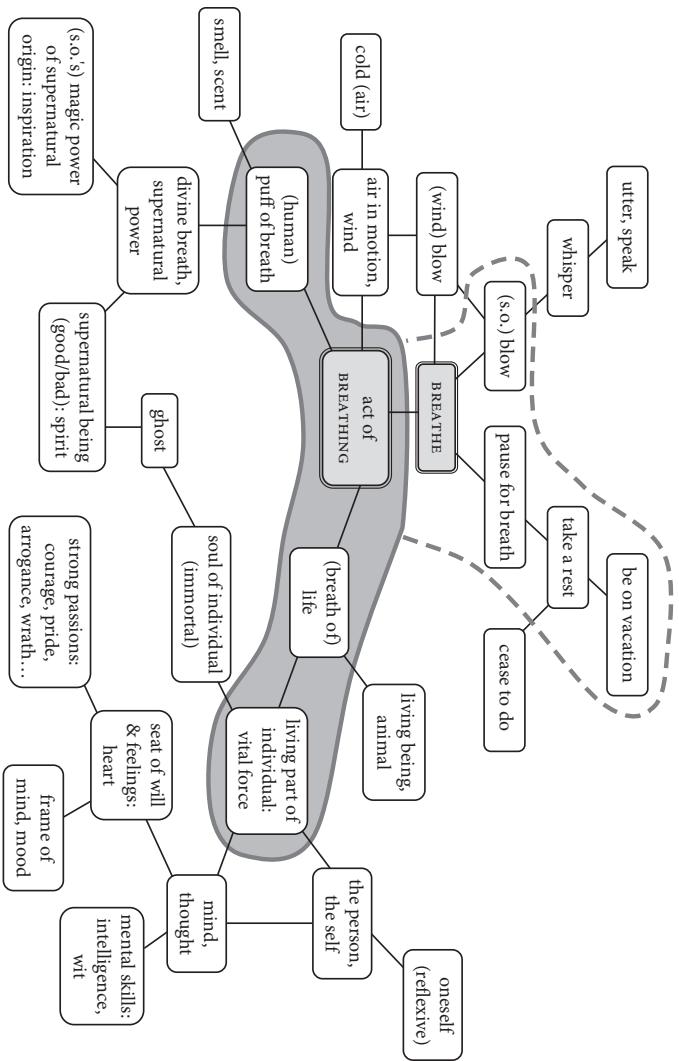
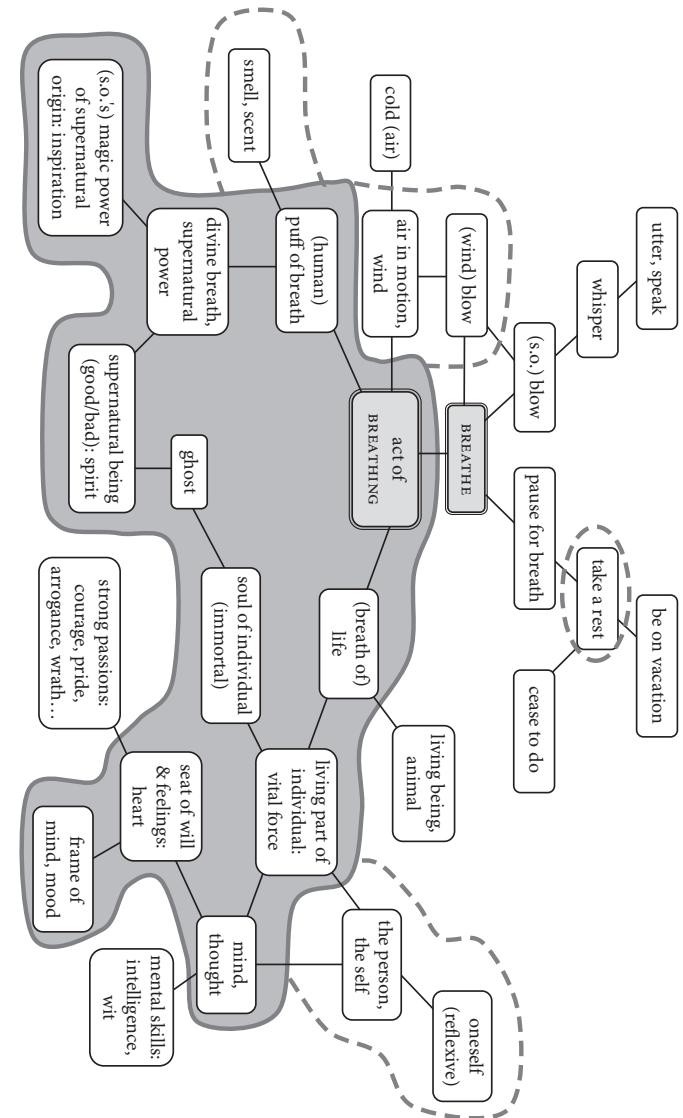
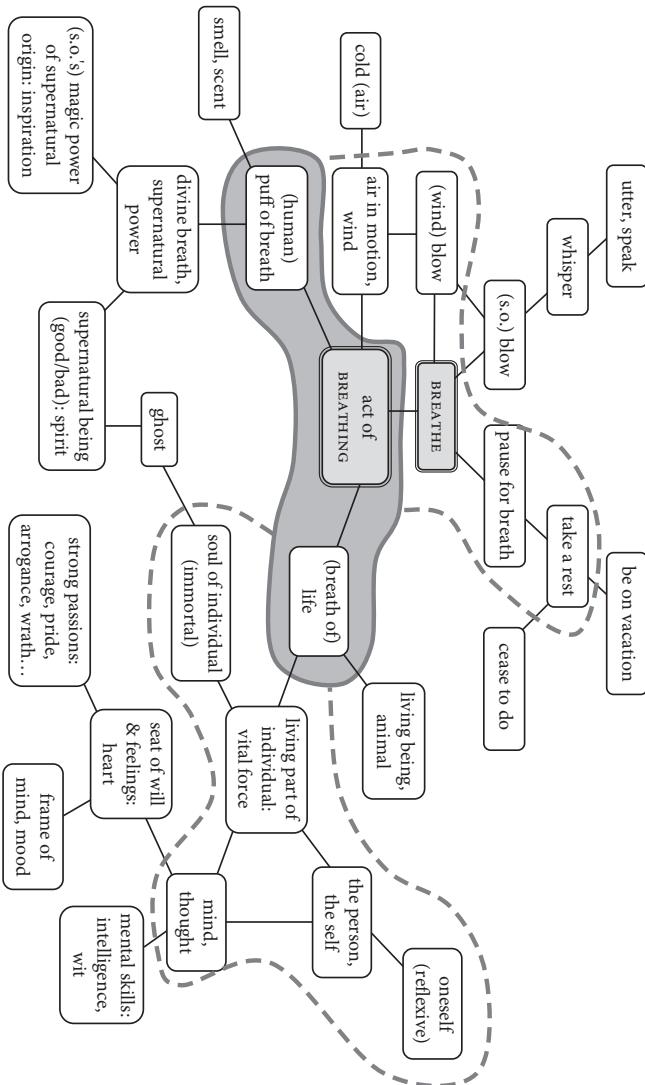


Figure 14. Isolectic set for Mwotlap mōkhe-.

Figure 15. Isolectic set for Standard Arabic *nūh*.Figure 16. Isolectic set for Standard Arabic *nafās*.

## References

- [no author] 1990. 汉语词典 *Hanyu Cidian: Dictionnaire Chinois-Français*. 1990. Paris: The Commercial Press.
- [no author] 1996[1978]. 现代汉语词典 *Contemporary Chinese Dictionary*. Beijing: Commercial Press.
- Anderson, L. 1982. "Perfect" as a universal and a language-particular category. In *Tense-Aspect: Between Semantics and Pragmatics*, P.J. Hopper (Ed.), 227–264. Amsterdam: John Benjamins.
- Bailly, A. 1950[1894]. *Dictionnaire Grec-Français*. Paris: Hachette.
- Bril, I. 2000. *Dictionnaire nélémwa-nixumwak français-anglais (Nouvelle-Calédonie)* [Langues et Cultures du Pacifique 14]. Paris: Peeters.
- Brown, C. 2005a. Hand and Arm. In Haspelmath et al. (Eds), 522–525.
- Brown, C. 2005b. Finger and Hand. In Haspelmath et al. (Eds), 526–529.
- Camden, W. 1979. Parallels in structure of lexicon and syntax between New Hebrides Bislama and the South Santo language spoken at Tangoa. In *Papers in Pidgin and Creole Linguistics 2* [A-57], 51–117. Canberra: Pacific Linguistics.
- Clark, R. n.d. North and Central Vanuatu: A comparative study. Computer files, University of Auckland.
- Croft, W. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: OUP.
- Croft, W., Bat-Zeev Shyldkrot, H. & Kemmer, S. 1987. Diachronic semantic processes in the middle voice. In *Papers from the 7th International Conference on Historical Linguistics*, A. Giacalone-Ramat, O. Carruba & G. Bernini (Eds), 179–192. Amsterdam: John Benjamins.
- Enfield, N. 2003. *Linguistic Epidemiology: Semantics and Grammar of Language Contact in Mainland Southeast Asia*. London: Routledge/Curzon.
- Evans, N. 1992. Multiple semiotic systems, hyperpolysemy, and the reconstruction of semantic change in Australian languages. In *Diachrony within Synchrony*, G. Kellerman & M. Morrissey (Eds), 475–508. Bern: Peter Lang.
- Evans, N. & Wilkins, D. 2000. In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language* 76: 546–592.
- Fortescue, M., Jacobson, S. & Kaplan, L. 1994. *Comparative Eskimo Dictionary, with Aleut Cognates* [Alaska Native Language Center Research Paper 9]. Fairbanks AK: University of Alaska.
- François, A. 2002. *Araki. A Disappearing Language of Vanuatu* [Pacific Linguistics 522]. Canberra: Australian National University.
- François, A. Forthcoming. Des valeurs en héritage: Les isomorphismes sémantiques et la reconstruction des langues. In *Typologie et comparatisme. Hommages à Alain Lemaréchal* (provisional title), I. Choi-Jonin, M. Duval & O. Soutet (Eds), [Orbis Supplementa]. Louvain: Peeters.
- François, A. In preparation. *Mwotlap-French-English dictionary*.
- Gaffiot, F. 1934. *Dictionnaire Latin-Français*. Paris: Hachette.
- Geeraerts, D. 1993. Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics* 4: 223–272.
- Haspelmath, M. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In *The New Psychology of Language*, Vol. 2, M. Tomasello (Ed.), 211–243. Mahwah NJ: Lawrence Erlbaum Associates.
- Haspelmath, M., Dryer, M., Gil, D. & Comrie B. (Eds), 2005. *The World Atlas of Language Structures*. Oxford: OUP.
- Jurafsky, D. 1996. Universal tendencies in the semantics of the diminutive. *Language* 72: 533–578.
- Kay, P. & Maffi, L.M. 2005. Colour terms. In Haspelmath et al. (Eds), 534–545.
- Lakoff, G. 1987. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago IL: University of Chicago Press.
- Matisoff, J. 1978. *Variational Semantics in Tibeto-Burman: The "Organic" Approach to Linguistic Comparison* [Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics VI]. Philadelphia PA: Institute for the Study of Human Issues.
- Monier-Williams, M. 1970[1899]. *A Sanskrit-English Dictionary, Etymologically and Philologically Arranged with Special Reference to Cognate Indo-European Languages*. Delhi: Motilal Banarsiidas.
- Naïm, S. 2007. L'énoncé réfléchi et les stratégies de réflexivation dans des variétés dialectales de l'arabe. In *L'énoncé réfléchi*, A. Rousseau, D. Roulland & D. Bottineau (Eds), 301–319. Rennes: Presse Universitaire de Rennes.
- Nerlich, B., Todd, Z., Herman, V. & Clarke, D. (Eds). 2003. *Polysemy: Flexible Patterns of Meaning in Mind and Language* [Trends in Linguistics]. Berlin: Mouton de Gruyter.
- Pauliat, P. 1991. *Dictionnaire Français-Russe, Russe-Français* [Mars]. Paris: Larousse.
- Reig, D. 1983. *As-Sabil, Dictionnaire Arabe-Français, Français-Arabe*. Paris: Larousse.
- Riemer, N. 2005. *The Semantics of Polysemy: Reading Meaning in English and Warlpiri* [Cognitive Linguistics Research]. Berlin: Mouton de Gruyter.
- Roper, E.M. 1928. *Tu Beđawi. An Elementary Handbook for the Use of Sudan Government Officials*. Hertford.
- Rosch, E. 1973. Natural categories. *Cognitive Psychology* 4: 328–350.
- Sakhno, S. 2005. *Les 100 racines essentielles du russe*. Paris: Ellipses.
- Sandra, D. & Rice, S. 1995. Network analysis of prepositional meaning: Mirroring whose mind – the linguist's or the language user's? *Cognitive linguistics* 6(1).
- Saussure, F. de. 1972[1916]. *Cours de linguistique générale*. Text edited and presented by T. de Mauro. Paris: Payot.
- Stchoupak, N., Nitti, L. & Renou, L. 1987[1932]. *Dictionnaire Sanskrit-Français* [Publications de l'Institut de Civilisation Indienne]. Paris: Maisonneuve.
- Tyler, A. & Evans, V. 2001. Reconsidering prepositional polysemy networks: The case of over. *Language* 77: 724–765. (Reprinted in Nerlich et al. 2003, 95–160).
- Vine, W.E. 1985. *Vine's Complete Expository Dictionary of Old and New Testament Words*. Nashville TN: Thomas Nelson.
- Wierzbicka, A. 1992. *Semantics, Culture, and Cognition: Universal Human Concepts in Culture-Specific Configurations*. Oxford: OUP.



# 6

## *The languages of Vanikoro: three lexicons and one grammar*

---

ALEXANDRE FRANÇOIS

### Discovering history through language: papers in honour of Malcolm Ross

Edited by Bethwyn Evans

Cette île, toute petite qu'elle est, présente le singulier phénomène de plusieurs idiomes differens. (Gaimard 1833:338)

#### 1 The paradox of Vanikoro languages<sup>1</sup>

##### 1.1 The languages of Vanikoro

With its 193 sq. km, Vanikoro is the second largest island in the small archipelago formerly known as the Santa Cruz Islands, and now often referred to as ‘Temotu’, after the official name of the easternmost province of the Solomon Islands (Map 1).

The province of Temotu is home to a variety of languages (Tryon 1994): three Polynesian, and nine non-Polynesian. The latter include three languages on Vanikoro, and three on Utupua — a total of six Oceanic languages which have long been understood to form a branch of their own (Tryon and Hackman 1983). The three remaining languages, known as the ‘Reefs-Santa Cruz’, were long deemed to be Papuan (Wurm 1976), but have recently been shown to be Austronesian (Ross and Næss 2007). More specifically, Ross and Næss have proposed to group all the non-Polynesian languages of the region into a single first-order subgroup of Oceanic, labelled ‘Temotu’. The latter would then split into two branches: Reefs-Santa Cruz (RSC) on the one hand, and Utupua-Vanikoro (UV)<sup>2</sup> on the other hand.

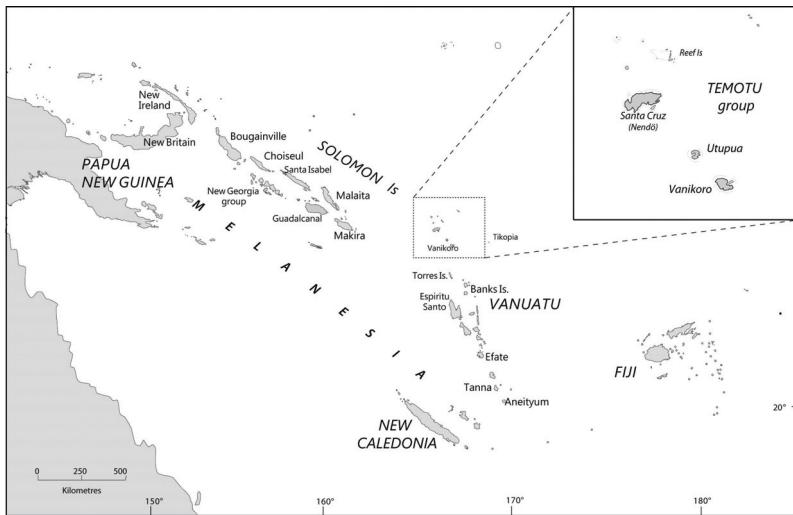
---

<sup>1</sup> My gratitude goes to Malcolm for having fostered my linguistic research on Melanesian languages, both through his writings and through our discussions, ever since we first met in 1997. His interest in my data has been a strong incentive for me to spend the years 2009 and 2010 at The Australian National University. My initial work on Vanikoro was facilitated by the Institut de Recherche pour le Développement of Nouméa and Association Salomon; by the French Centre National de la Recherche Scientifique; by Piet Lincoln; by Association ‘Banie’ and the traditional chiefs of the island; and by my Vanikoro helpers and friends — especially Stanley Repuamu, Ezekiel Prians, John Nabu. I am grateful to Bethwyn Evans and Andrew Pawley for their comments on earlier versions of this paper.

<sup>2</sup> Except for its new location in the POC tree, this UV branch coincides with the subgroup identified earlier as ‘Eastern Outer Islands’ (Tryon 1994, 1995).



Pacific Linguistics  
Research School of Pacific and Asian Studies  
The Australian National University



Map 1: Location of Vanikoro in the Pacific

If Ross and Næss' hypotheses are correct, then Temotu constitutes a new branch of the Oceanic family tree whose history needs to be written. The present paper hopes to play its part in this endeavour, by presenting primary data and some discussion on the three indigenous languages of Vanikoro: Teanu (TEA), Lovono (LVN) and Tanema (TNM).

Published information on Vanikoro languages is still limited, but certainly not absent. In 1788, the island of Vanikoro was reached by the two frigates of the French navigator Jean-François de La Pérouse, and saw his fatal shipwreck — an event which was only understood a few decades later by the Irish navigator Peter Dillon (Dillon 1830). Another French officer, Dumont d'Urville, immediately organised an expedition in Dillon's wake. Among the abundant documentation produced by this second voyage (Dumont d'Urville 1830–1834), the French naturalist Gaimard (1833, 1834) compiled wordlists in the three languages of Vanikoro, a precious document on the linguistic situation of that time.<sup>3</sup> Much more recently, other word lists have been compiled by Tryon and Hackman (1983), based on an extended version of the Swadesh basic vocabulary list. Darrell Tryon also wrote short grammatical accounts of Lovono (Tryon 1994:630–634), and of Teanu (Tryon 2002). Additionally, a short collection of Teanu texts was published in Tua and Lincoln (1979).<sup>4</sup>

Almost two centuries after Dumont d'Urville's expedition, the French Ministère de la Marine, together with Association Salomon and Institut de Recherches pour le Développement, organised another expedition called *Vanikoro 2005*, to find out about the fate of La Pérouse's ships and sailors. I was given the opportunity to play my part there as

<sup>3</sup> I am much indebted to Piet Lincoln for allowing me easy access to Gaimard's documents.

<sup>4</sup> The two languages Teanu and Lovono have been given varying names over time. Teanu was called *Tanéanou* by Gaimard, and *Buma* by Tryon. Lovono was called *Vanikoro* by Gaimard, *Vanikolo* by Ivens (1918), *Vano* by Tryon. See §1.2 below for a discussion of my naming proposals.

a linguist, documenting place-names and oral traditions, with a special interest in the islanders' stories that still remember so vividly the 1788 wreckage (François 2008a). On this occasion, I was also able to follow in Gaimard's footsteps, and record what I could of the three languages. One thing I realised was the urgency of this task, with both Lovono and Tanema remembered by only a handful of speakers.

## 1.2 A note on the history and geography of Vanikoro

The population of Vanikoro can be described at two different levels of observation.

A contemporary look would probably suggest just a binary divide between two communities, one Melanesian and one Polynesian. The latter is a group of about 300 settlers originating from Tikopia, a small island located about 200 km eastwards. Although they have been colonising the southern shores of Vanikoro for more than three centuries (see Dillon 1830), they tend to interact very little with the native population — except for the occasional land dispute. As their social network is still anchored in their Polynesian homeland, they remain predominantly monolingual in Tikopian, the Polynesian Outlier language spoken on Tikopia (Firth 1985). Apart from a few loanwords here and there, this recent colonisation does not show any major linguistic consequence, and will not be discussed further.

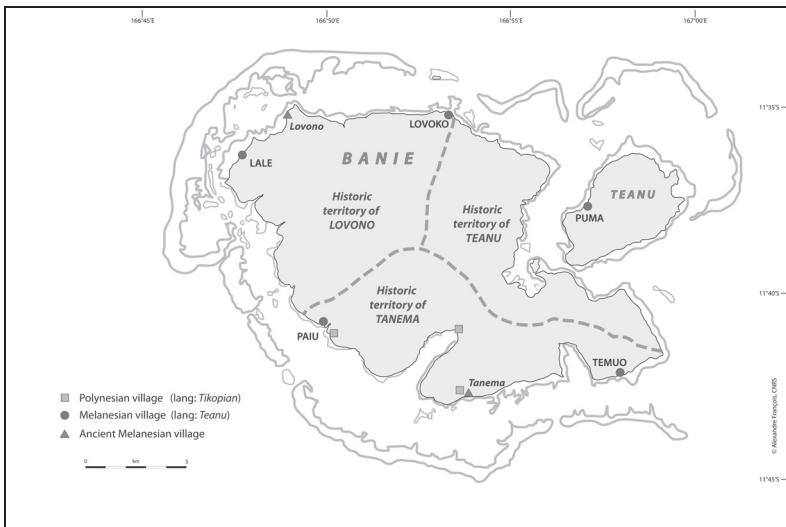
With about 600 individuals descending from the earlier inhabitants of Vanikoro, the Melanesians live today in six coastal villages scattered around the island: Puma, Temuo, Paiu, Lale, Lovono, Lovoko. People from these villages communicate through coastal canoeing, and now form a homogeneous society. This modern unity reportedly results from the action of the Anglican missionaries who christianised the region at the end of the 19<sup>th</sup> century. However, the first historical documents on Vanikoro (Dillon 1830; Dumont d'Urville 1830–1834), as well as the oral tradition of the islanders themselves, tell a different story: that of an island which used to be sharply divided into three distinct 'tribes' or chiefdoms. Each of these tribes was attached to a specific territory which they defended fiercely from one another, and which is still clearly delimited in people's memories, see Map 2.

Each tribe's name recalls a significant place of its own area:

- **Teanu** from the name of the northeast island of the Vanikoro group, where the village of *Puma* is also located<sup>5</sup>
- **Lovono** from the name of a village, also formerly known as *Vono* or *Vano*,<sup>6</sup> northwest of the main island Banie
- **Tanema** an ancient village, also known as *Tetawo*, on the southern coast (in what is now *de facto* Polynesian territory)

<sup>5</sup> The village name *Puma*, sometimes with the incorrect spelling *Buma*, has been used as another name for the language of Teanu (see footnote 4).

<sup>6</sup> The Lovono name of this village of Vano/Lovono was apparently *Vana* or *Alavana*. In principle, one might want to use this autonym as a reference name for this language, yet this would be slightly artificial: the only term which is used today, even by the last speakers themselves, is the form *Lovono* in Teanu, the only surviving language.



Map 2: The populations of Vanikoro

Each of these three tribes had its own language, which survived up until the 20<sup>th</sup> century. Gaimard's (1833, 1834) wordlists cite 'Tanéma' and 'Vanikoro' (Lovono) on an equal footing with 'Tanéanou' (Teanu). Ivens (1918:155), in his translations of the prayer 'Our Father', illustrates Vanikoro with a language that can be identified as Lovono — which suggests it was still in a healthy state at the end of the 19<sup>th</sup> century. But since the pacification of the island, intermarriage amongst the three earlier communities has increased, and they have merged into a single society. Probably due to its earlier demographic lead, Teanu was adopted as the whole island's daily language, very much at the expense of the two other vernaculars. In 2005, Lovono and Tanema were still remembered by only a handful of elder speakers — five for Lovono, four for Tanema. Because these last speakers are now dispersed across the various Teanu-speaking villages of Vanikoro, they do not form anything that would resemble a speech community. Clearly, the shift to Teanu is now complete: Lovono and Tanema are now two moribund languages, with only a few years left to live.

### 1.3 Three closely related languages

#### 1.3.3 Three aberrant Oceanic languages

For the linguist accustomed to other Oceanic-speaking areas, the three languages of Vanikoro can be disconcerting. This section will touch briefly upon the question of their genetic affiliation, before we examine the issue of their mutual relationship.

One conspicuous characteristic of Teanu, Lovono and Tanema is the degree to which they differ, both lexically and morphologically, from most other known Oceanic languages (François 2006), whether from the Solomons or from nearby Vanuatu. For example, these

three languages show relatively little retention of the lexicon reconstructed for Proto Oceanic. Many etyma, which tend to be otherwise widely preserved among Oceanic languages (Pawley 2007), have disappeared altogether from the lexicon of Vanikoro languages; e.g. \*tama- 'father', \*mate 'die', \*kani 'eat', \*kayu 'tree', \*ikan 'fish', \*sapa 'what', \*jalan 'road', \*susu 'milk', \*pano 'go', \*sake 'up', \*sipo 'down'. The morphology also shows a great deal of erosion. Thus the languages of Vanikoro show no trace whatsoever of the POc possessive affixes (\*-gu, \*-mu, \*-ña...; \*-qi...), of object pronouns, of the article \*na, of common verbal affixes like the transitive suffix \*-i, the applicative \*-aki[n], the causative \*pa[ka]-, and so on.

In sum, Vanikoro languages can be considered 'aberrant' (see Grace 1990; Pawley 2006) in comparison with most other Oceanic languages. This does not mean, however, that their Oceanic ancestry cannot be detected. Indeed, despite the high level of lexical replacement which evidently took place in their history, it is still possible to uncover some words whose similarity with Oceanic reconstructions is beyond doubt, and which could hardly be attributed to chance or borrowing — if only because their phonological correspondence patterns seem to follow some degree of regularity. Table 1 provides a small sample list of such words.<sup>7</sup>

Table 1: Some obvious Oceanic reflexes across the three languages

English	Teanu	Lovono	Tanema	POc etymon
'bird'	menuko	menuka	manuke	*manuk
'eye'	mata	mala	ka <sup>v</sup> mae	*mata
'soul, spirit'	ata	ala	ae	*qata
'ears'	taña	mabe <sup>v</sup> leje	añe	*taliqa
'house'	moe ~ mwœ	moe	nalama	*rumaq
'cold'	medigo	menija	medija	*ma <sup>(n)</sup> ri <sup>(n)</sup> rij
'long, tall'	biouro	beure	va <sup>v</sup> beura	*barapu
'Canarium nut'	vo <sup>v</sup> joro	ve <sup>v</sup> jere	vi <sup>v</sup> jara	*[ka]jari
'lie down'	wene	enu	eno	*qenop

#### 1.3.2 Proto Vanikoro, the common ancestor

Not only can Teanu, Lovono and Tanema be individually linked to Proto Oceanic, but they can also be shown to form a set of three closely related languages. It is possible to suggest instances of exclusively shared innovations, thereby pointing to the likely existence of a common ancestor Proto Vanikoro. The following paragraphs therefore answer the question raised by Ross and Næss (2007:473), according to whom 'no innovations define Vanikoro'.

<sup>7</sup> The three Vanikoro languages have the same phoneme inventory, with five short vowels (i, e, a, o, u) and 19 consonants. The spelling conventions here adopted include v=/β/; ŋ=/ŋv/; j=/<sup>v</sup>j/. Labiovelar consonants use digraphs with a w, and all prenasalised voiced stops are spelled without their nasal element, e.g. /m<sup>v</sup>b<sup>v</sup>/ is bw, /p<sup>v</sup>g<sup>v</sup>/ is g, etc.

In the phonological domain, one can cite the addition of a paragogic vowel, probably a schwa, after most word-final consonants, resulting in the retention of that consonant followed by a non-etymological vowel:

- (1) POc \*manuk ‘bird’ > TEA *menuko* ~ LVN *menuka* ~ TNM *manuke*;  
POc \*ma<sup>n</sup>riŋriŋ ‘cold’ > \*ma<sup>n</sup>riŋja > TEA *medigo* ~ LVN *meniŋe* ~ TNM *medija*;  
POc \*wair ‘water’ > \*waiřa > TEA *ero* ~ LVN *wire* ~ TNM *n'ira*.

Another example of a sound change which is only found in Vanikoro languages is the occasional velarisation of \*t to /k/ before a high back vowel /u/:

- (2) POc \*patu ‘stone’ > \*vatu > \*vakə > TEA *voko* ~ LVN/TNM *vaka*;  
POc \*kutu ‘louse’ > \*utu > \*uku > \*ukə > TEA *uko* ~ LVN/TNM *-uka*.

In the morphological domain, we will see (§3.1.2) that the three languages share the same structural collapse between certain non-singular personal prefixes.

Finally, many lexical items seem to be shared exclusively by these three languages. Table 2 proposes a set of possible lexical innovations in Proto Vanikoro — with tentative reconstructions of the most likely etymological forms.<sup>8</sup>

**Table 2:** Some putative lexical innovations in Proto Vanikoro

English	Teanu	Lovono	Tanema	Proto Vanikoro
‘moon’	<i>metele</i>	<i>mele</i>	<i>maloula</i>	*mataul(ə)
‘fish’	<i>namuko</i>	<i>namweka</i>	<i>namaka</i>	*nam <sup>(w)</sup> əkə
‘wood, tree’	<i>vilo</i>	<i>kuile</i>	<i>veila</i>	*v <sup>(w)</sup> eilə
‘taro’	<i>je'bute</i>	<i>bule</i>	<i>bue</i>	*bute
‘child’	<i>menu</i>	<i>melika</i>	<i>anuka</i>	*menuk(ə)
‘woman’	<i>emele</i>	<i>neme</i>	<i>me</i>	*nemel(ə)
‘name’	<i>eŋa</i>	<i>neŋe</i>	<i>niŋe</i>	*neŋə
‘who’	<i>ŋele</i>	<i>eŋe</i>	<i>ŋela</i>	*eŋel(ə)
‘inhabited land’	<i>kulumoe</i>	<i>kulamoe</i>	<i>kulama</i>	*kuləma? (ə)
‘rubbish’	<i>ajekele</i>	<i>togale</i>	<i>soge</i>	*jogel(ə)
‘tell s.o.’	<i>viňi</i>	<i>puňi</i>	<i>puňa</i>	*puňə
‘heavy’	<i>motoro</i>	<i>melure</i>	<i>mwaura</i>	*matur(ə)
‘dirty’	<i>sukiro</i>	<i>sukure</i>	<i>vatukura</i>	*sukir(ə)
‘stand’	<i>vio</i>	<i>pia</i>	<i>veo</i>	*piə
‘impede, protect’	<i>botojo</i>	<i>beloŋa</i>	<i>baoŋa</i>	*batonŋə
‘do again’	<i>tabo</i>	<i>lebu</i>	<i>abo</i>	*tabo

<sup>8</sup> Due to the lack of solid regularity, in particular, in the domain of vowels, it is difficult to securely reconstruct any protoform based on the synchronic data. This is an important difference between the languages of Vanikoro and other parts of Oceania, where protoforms can be reconstructed based on modern reflexes. See Ross (1988) for the languages of western Melanesia, Lynch (2001) for southern Vanuatu, François (2005) for northern Vanuatu.

In order to ascertain these reconstructions and expand the list of Proto Vanikoro innovations, more needs to be known of the languages on the neighbouring island Utupua. But the evidence given here should be sufficient to establish that the three Vanikoro languages form a subgroup of their own, pointing to a single common ancestor.

#### 1.4 Divergence and convergence among Vanikoro languages

In sum, Teanu, Lovono and Tanema are three ‘aberrant’ Oceanic languages which are genetically closely related to each other, as descendants of a single ancestor. Now as we compare the three modern languages, we may be surprised by a paradox:

- Compared to the situation in most other Oceanic languages, the forms of words in the three languages of Vanikoro tend to differ from each other in ways which can be regarded as quite extreme given their geographic closeness.
- Despite their heterogeneity with regard to word forms, they show perfect isomorphism of their structures.

This paradox will be the main focus of this article.

Throughout this paper, I will refer to two essential components of language, which crosscut the traditional division between lexicon and grammar. On the one hand, I will mention STRUCTURES, referring to the various concepts and semantic categories with which a language divides up semantic space — whether this refers to lexical or to grammatical meaning. On the other hand, each language embodies these categories and concepts into linguistic FORMS, endowed with a specific phonological content.<sup>9</sup> Two languages can be said to be isomorphic in a particular domain of their system, if they share the same structures or semantic categories, whether or not the forms they use are cognate with each other. For example, the two constructions *I have caught cold* and *J'ai attrapé froid* are perfectly isomorphic, because they express the same event by resorting to exactly identical metaphors and categories.

My observation is that the three Vanikoro languages exhibit a high degree of divergence in their forms, yet still show an extreme isomorphism of their structures. This configuration is illustrated in example (3). As far as the grammar is concerned, the three languages possess parallel structures and word order, to the point that they can all be analysed with a single line of word-to-word glosses. Yet on the other hand, one can equally note the dissimilarity between the actual forms of their words.<sup>10</sup>

- (3) TEA    A-ko               u-ka               u-katau               ene ?  
      LVN    Nu-pu             ku-ma             ku-ki                yane ?  
      TNM    Go-po             go-loma           go-ie               nana ?  
            2SG:R-say    2SG:IR-come    2SG:IR-follow    1SG  
            ‘Do you want to come with me?’

<sup>9</sup> This contrast STRUCTURES versus FORMS corresponds to what Hjelmslev (1961:52) described as respectively *content form* versus *expression form*; and to what Grace (1981:24) would call *content form* versus *lexification*.

<sup>10</sup> Abbreviations in glosses include: R - realis prefix; IR - irrealis prefix; GEN - General possessive classifier; INDEP - independent pronoun; FOOD - possessive classifier for food possession; HUM - article for human referents.

This observation, whereby languages can be at once homogeneous in structure while dissimilar in forms, has been widely made in the linguistic literature already (see Gumperz 1971; Enfield 2001). However, most of the time, those facts of structural parallelism result historically from contact between genetically diverse languages. To focus on works on the Melanesian area, Thurston (1989, 1994) thus describes the structural similarities between languages of northwestern New Britain, some Papuan and some Oceanic, and Ross (1996, 2001) discusses the influence of Waskia (Papuan) upon Takia (Oceanic), in Papua New Guinea's Madang Province. To this list, one could add discussions of structural parallelism between the English-based pidgins of the Pacific and their Melanesian substrates (Camden 1979; Keesing 1988, 1991; Siegel 2008). In all these cases, whether one compares Oceanic with Papuan or with European languages, the diversity of forms is a given; and what is observed is first and foremost a matter of STRUCTURAL CONVERGENCE — or ‘metatypy’, to use the term coined by Malcolm Ross (1996, 2001).

But the case of Teanu, Lovono and Tanema raises different issues, because they belong to the same genetic subgroup. For such closely related languages to share the same syntactic structures may partly reflect the mere legacy of their common ancestor, and partly be explained by later contact-induced convergence. The intriguing part here is rather the dissimilarity of forms: it needs to be considered not as a simple given — as was the case for genetically diverse languages — but as the problematic result of historical DIVERGENCE from a common ancestor. This configuration therefore requires specific explanations, beyond the now well-known cases of metatypy.<sup>11</sup>

Section 2 will discuss the degree of similarity and dissimilarity existing between the lexical forms of Vanikoro languages. Section 3, in turn, will demonstrate their strong structural isomorphism. Section 4 will finally propose a functional hypothesis to account for this linguistic paradox of Vanikoro languages.

## 2 Similarity and divergence of lexical forms

The linguistic relations between the three Melanesian languages of Vanikoro can thus be summarised in a simple formula: DISSIMILAR FORMS, SIMILAR STRUCTURES. The present §2 will discuss the first of these two dimensions, by assessing the degree of differentiation between the lexicons of Teanu, Lovono, and Tanema.

Even though Teanu, Lovono and Tanema are close genetic relatives (§1.2), the impression that prevails is that of a rather strong dissimilarity of their lexicons. Indeed, contrary to the impression given by Table 1 above, the three languages of Vanikoro are not mutually intelligible. Their dissimilarities are of varying nature, going from more or less regular phonological differences between cognate words, to forms that are simply non-cognate. Overall, while such formal differences are frequent in the Melanesian area, their degree is here rather impressive for languages which are spoken on the same island, and appear to have historically differentiated on this island.

<sup>11</sup> A similar blend of formal dissimilarity and structural parallelism can be found in the languages of north Vanuatu (François 2007, forthcoming, in prep.). However, the phenomenon appears to be even more conspicuous in the case of Vanikoro.

### 2.1 Differences due to phonological change

In some cases, forms which are superficially very dissimilar can in fact be explained by regular processes of sound change. While many correspondences between the three languages of Vanikoro are straightforward and obvious, some are more drastic and can result in little resemblance between the actual forms.<sup>12</sup>

For example, the verbs for ‘sit, stay’ (TEA *te* ~ LVN *lu* ~ TNM *o*) are dissimilar enough to suggest they might be non-cognate. However, one can establish a regular correspondence pattern TEA /t/ ~ LVN /l/ ~ TNM Ø (see sample in Table 3) pointing to a proto-consonant \**t* (Tryon and Hackman 1983:71).

Table 3: Some lenis reflexes of POc \**t*

English	Teanu	Lovono	Tanema	POc etymon
‘sit, stay’	<i>te</i>	<i>lu</i>	<i>o</i>	* <i>toka</i>
‘three’	<i>te-te</i>	<i>te-lu</i>	<i>a-o</i>	* <i>tolu</i>
‘sugarcane’	<i>to</i>	<i>lepie</i>	<i>ova</i>	* <i>topu</i>
‘soul, spirit’	<i>ata</i>	<i>ala</i>	<i>ae</i>	* <i>qata</i>
‘ghost, spirit’	<i>tadoe</i>	<i>leñoe</i>	<i>aoe</i>	* <i>qata-</i> <sup>?</sup>
‘do again’	<i>tabo</i>	<i>lebu</i>	<i>abo</i>	
‘carry on shoulders’	<i>tabe</i>	<i>lebe</i>	<i>ebe</i>	
‘impede, protect’	<i>botoño</i>	<i>belona</i>	<i>baona</i>	
‘unripe, new’	<i>motoe</i>	<i>meloe</i>	<i>maja</i>	* <i>mataq</i>

The three forms for ‘sit, stay’ therefore suggest an etymon \**tV*. While correspondences are much less obvious regarding vowels, a proto-form with /o/ is a likely origin for a pattern TEA /e/ ~ LVN /u/ ~ TNM /o/; see also the vowels of ‘lie down’ (< POc \**qenop*) in Table 1, and of ‘three’ in Table 3. The perfectly parallel reflexes for ‘three’ (< \**to* < POc \**tolu*) and for ‘sit, stay’ confirm a reconstruction \**to* — certainly the first syllable of POc \**toka* ‘stay’. Interestingly, Gaimard (1833, 1834) consistently writes these Tanema words with an *r*, which shows the correspondence pattern was TEA /t/ ~ LVN /l/ ~ TNM /r/ two centuries ago. Thus, he writes *rarou* for ‘three’ (modern *a-o*), and *guidiro* for what he glosses ‘Asseyez-vous’; the latter in fact representing \**giti-ro* (modern *giti-o*) ‘we[INCL] are sitting’.

In sum, the regularity of correspondences, when they can be established, makes it possible, quite classically, to detect the cognacy of some forms which would have otherwise seemed unrelated.

Sometimes, the ultimate POc source of a given series is unclear, yet at least one can tentatively draw connections between modern forms, based on synchronic regular sound correspondences. For example, the verb for ‘die’ is *bu* in Teanu and Tanema, and *me* in Lovono, two forms with little in common. However, a regular sound pattern seems once again to emerge from the data: TEA /b/ ~ LVN /m/ ~ TNM /b/ — with a small amount of variation involving voicing or rounding of the consonant (Table 4).

<sup>12</sup> Some of the regular correspondences are given in Ross and Næss (2007).

**Table 4:** A possible regular correspondence pattern

English	Teanu	Lovono	Tanema	POc etymon
‘die, dead’	<i>bu</i>	<i>me</i>	<i>bu</i>	(*mate) ?
‘sharpen, trim’	<i>bo</i>	<i>me</i>	<i>bo</i>	
‘1exc:dual pronoun’	<i>keba</i>	<i>gema</i>	<i>gabe</i>	*kama[m]ji
‘k.o. basket’	<i>iunubo</i>	<i>nunumie</i>	<i>nuba</i>	
‘tobacco’	<i>nabene</i>	<i>nakamene</i>	<i>nabwane</i>	
‘blood’	<i>abo</i>	<i>amwale</i>	<i>aba</i>	
‘Areca catechu’	<i>buioe</i>	<i>namwe</i>	<i>buaia</i>	*buaq
‘Reflexive-Reciprocal’	<i>ñepe</i>	<i>ñeme</i>	<i>be</i>	
‘fishing net’	<i>pele</i>	<i>menele</i>	<i>benala</i>	*kup <sup>w</sup> ena ?

This suggests the three forms for ‘die’ may be cognate, despite their present dissimilarity. In this case, because no known POc reconstruction (including \*mate) provides any satisfying etymon, the cognacy judgment rests on purely synchronic data.

**Table 5:** Some potential, but dubious, cognate sets

English	Teanu	Lovono	Tanema	POc etymon
‘man, person’	<i>mwaliko</i>	<i>lamuka</i>	<i>anuka</i>	*m <sup>w</sup> aqane ?
‘women’	<i>viñevi</i>	<i>venime</i>	?	*pine ?
‘canoe, ship’	<i>kuo</i>	<i>nawe</i>	<i>goia</i>	*waga(ŋ) ?
‘sleep’	<i>mokoiu</i>	<i>mepeu</i>	<i>matou</i>	*maturur ?
‘red’	<i>moloe</i>	<i>wamoene</i>	<i>manobeila</i>	*meraq ?
‘randomly; in vain’	<i>moli</i>	<i>moli</i>	<i>mano</i>	
‘help; with’	<i>samame</i>	<i>emeio</i>	<i>avaio</i>	
‘where?’	<i>vele</i>	<i>mane</i>	<i>vane</i>	
‘I, 1sg pronoun’	<i>ene</i>	<i>yane</i>	<i>nana</i>	
‘sink’	<i>metelu</i>	<i>mwelesu</i>	<i>madilo</i>	
‘perhaps’	<i>bwara</i>	<i>bweti</i>	<i>buru</i>	
‘go down’	<i>abu</i>	<i>pwo</i>	<i>kabu</i>	
‘good’	<i>wako</i>	<i>vakane</i>	<i>apika</i>	
‘seize, hold’	<i>labu</i>	<i>lo</i>	<i>nou</i>	
‘rejoice’	<i>pei</i>	<i>pwadi</i>	<i>pae</i>	
‘stone oven’	<i>awene</i>	<i>epene</i>	<i>pavene</i>	

Occasionally, the modern forms display little more than a vague ‘family resemblance’. That is, the modern lexical forms are possibly cognate, yet their phonemes enter no regular correspondence pattern, in such a way that one could only explain their cognacy by resorting to *ad hoc* etymological hypotheses. Table 5 provides a sample of such potential, but irregular and dubious, cognate sets, which would require closer scrutiny in the future.

## 2.2 Lexical replacement

Finally, it also often happens that the three languages have forms that are distinct, and almost certainly not cognate for the same meaning. A sample of such cases is given in Table 6.

**Table 6:** Some non-cognate sets resulting from lexical innovation

English	Teanu	Lovono	Tanema
‘thing’	<i>yatene</i>	<i>vesemele</i>	<i>vamora</i>
‘know’	<i>ovei</i>	<i>lonei</i>	<i>wo</i>
‘come’	<i>ka</i>	<i>mage</i>	<i>loma</i>
‘lie, deceive’	<i>tomoli</i>	<i>ñaine</i>	<i>role</i>
‘see’	<i>romo</i>	<i>eti</i>	<i>rungi</i>
‘quickly’	<i>kiane</i>	<i>segei</i>	<i>gamoi</i>
‘bad’	<i>tamwaliko</i>	<i>visale</i>	<i>vae</i>
‘big (PLUR)’	<i>wopine</i>	<i>evala</i>	<i>bwau</i>
‘broken’	<i>mamakoe</i>	<i>seli</i>	<i>vave</i>
‘remote’	<i>somu</i>	<i>akaole</i>	<i>mosomu</i>
‘down, below’	<i>puo</i>	<i>lenu</i>	<i>ese</i>
‘what?’	( <i>yan</i> ) <i>ae</i>	<i>ese</i>	<i>sive</i>
‘do what, do how?’	( <i>mi</i> ) <i>kae</i>	<i>ñese</i>	<i>jive</i>
‘be why?’	<i>ve</i>	<i>wo</i>	<i>ja</i>
‘another, an’	<i>iote</i>	<i>leka</i>	<i>keo</i>
‘one; same’	<i>iune</i>	<i>tilioko</i>	<i>omwano</i>
‘play; wander’	<i>moloe</i>	<i>telu</i>	<i>lumota</i>
‘neck; mind’	<i>awa</i>	<i>warene</i>	<i>vasare</i>
‘bush’	<i>yogoro</i>	<i>atera</i>	<i>arara</i>
‘cabbage’	<i>tebo</i>	<i>lamware</i>	<i>some</i>
‘rat’	<i>uvilo</i>	<i>katone</i>	<i>ivala</i>
‘be ripe’	<i>ako</i>	<i>wi</i>	<i>kou</i>
‘guts’	<i>bea</i>	<i>lale</i>	<i>lebwe</i>
‘year’	<i>ebieve</i>	<i>verue</i>	<i>rove</i>
‘chicken’	<i>kulevelu</i>	<i>kio</i>	<i>tokila</i>
‘leaf’	<i>uie</i>	<i>nugia</i>	<i>lele</i>

In their survey of Solomon Island languages, Tryon and Hackman (1983:481) give the following cognate percentages for the three languages of Vanikoro, based on a modified version of Swadesh’s basic vocabulary list (200 words):

- Teanu–Lovono 57.3 %
- Teanu–Tanema 51.1 %
- Lovono–Tanema 54.8 %

These figures point to a relatively high level of lexical dissimilarity. They surpass similar counts made in most other parts of island Melanesia. For example, the Torres and Banks Is

of north Vanuatu are another area where lexical replacement has been intense (François, in prep.). However, in order to find figures as low as those on Vanikoro, one has to pick languages which are geographically spread apart; e.g. 44.5 % between Hiw (Torres) and Lakon (Gaua, south Banks). Should one consider languages spoken on a single island, the widest gap one can find there is between Lakon and Dorig, with 61.5 % shared vocabulary (Tryon 1976:95).

Even more instructive is the comparison with other language families in the world. To take just one example, rates of shared vocabulary amongst Germanic languages do not go below 53.6% (Dyen, Kruskal and Black 1992). In other words, the three languages of Vanikoro have managed to achieve, within the limited space of a single island, more lexical diversity than the whole Germanic subgroup of Indo-European. Such an intense differentiation deserves to be acknowledged, and discussed (see §4).

The strong formal dissimilarity that prevails between the three languages of Vanikoro naturally results in sentences where the phonological form of words, whether lexical or grammatical, can show a high degree of dissimilarity:

- (4) TEA *Pi-te ne sekele iupa, pi-wowu uo.*  
 LVN *Nupe-lu ne amenoya iemitore, nupe-ŋoa upie.*  
 TNM *Tei-o ini vasayola akegamuto, ti-oa uva.*  
 1EX.PL:R-stay in garden our 1EX.PL:R-plant yam  
 'We were in our garden, we've been planting yams.'
- (5) TEA *Somu tamwase tae, vitoko takoie ne.*  
 LVN *Akaole visalewabeu taie, vateoko yate ida.*  
 TNM *Mosomu vaepamabo eia, vatako eto kana.*  
 remote very NEG close uphill here  
 'It is not very far; it's close to here, up this way.'

The impression of formal heterogeneity may be due partly to sound change affecting words that are in fact cognate (e.g. 'yam', 'stay', 'plant'); and partly to lexical replacement (e.g. forms for 'very', 'remote', 'uphill') or morphological change (e.g. forms of personal pronouns).

What is perhaps more puzzling is the contrast between, on the one hand, this formal dissimilarity, and on the other hand, the perfect parallelism existing between these languages' structural properties. This is the topic of the next section.

### 3 Structural isomorphism

The push towards linguistic differentiation has affected the phonological forms of words, yet evidently had little impact on grammatical and semantic structures. Even as their lexicons were diverging from each other, the three languages of Vanikoro have maintained a strong STRUCTURAL ISOMORPHISM — whether in syntax, phraseology or organisation of meaning. This is reflected, for instance, by the ability of translating word-for-word any sentence from one language to another, following the same word order and the same morphological and lexical categories — as in examples (3) to (5).

The present section will illustrate this strong structural parallelism using a few conspicuous examples from various aspects of the grammar, and will end with a tentative account of the linguistic history of Vanikoro.

### 3.1 Morphosyntax

#### 3.1.1 Syntax of the clause

Teanu, Lovono and Tanema are parallel in all aspects of their syntax. This includes all properties related to word order, whether the basic SVO clause order, or other properties, use of prepositions, post-nominal placement of adjectives and possessors, clause-final position of the negation as in (5).

The three languages display the same organisation in parts of speech, including a sharp divide between nouns and verbs, and a distinctive category of adjectives. While all verbs are obligatorily prefixed for subject and mood (§3.1.2), adjectives are unprefixed. Languages even agree on whether they treat a specific notion as an adjective or as a verb, as evidenced by the parallel presence vs absence of the subject prefix in the two predicates of (6).

- (6) TEA *Udo ponu, boro we i-ako ?*  
 LVN *Puya pae, bware we i-wi ?*  
 TNM *Uda pade, betika we i-kou ?*  
 banana that black/unripe or 3SG:R-be.ripe  
 'Those bananas, are they green<sub>[ADJ]</sub> or ripe<sub>[VERB]</sub>?'

Likewise, all syntactic properties of the clause are reflected alike across the three languages. They all lack noun articles, case markers, verb transitivisers or applicatives, and causative affixes. They make use of a reflexive marker (Table 4), which is also a reciprocal and an emphatic marker. They all resort frequently to core-layer verb serialisation, with exactly parallel phrasing; see examples (3), (7) and (10). They possess exactly parallel Tense-Aspect-Mood categories, and so on.

#### 3.1.2 Pronouns and TAM marking

The paradigms of personal pronouns are also organised in parallel ways. Like most Oceanic languages, those of Vanikoro distinguish between exclusive and inclusive 'we'; and they show three numbers: singular, dual, plural.

Table 7 shows the eleven independent pronouns for the three languages. These show a reasonable degree of similarity.

Table 7: Personal independent pronouns

	Teanu	Lovono	Tanema
1sg	ene	ŋane	nana
2sg	eo	ago	go
3sg	ini	ŋani	nini
1in:du	kia	gita	gie
1ex:du	keba	gema	gabe
2du	kela	gamila	gamine
3du	da	dea	delalu
1in:pl	kiapa	giitu	geto
1ex:pl	kupa	gamitu	gamuto
2pl	kaipa	gaipa	gamito
3pl	dapa	detu	dato

Besides these free pronouns, these languages also possess<sup>13</sup> a double set of mood-marked subject prefixes for verbs (one for realis, one for irrealis); see Table 8.

**Table 8:** Verbal prefixes for subjects

	Teanu		Lovono		Tanema	
	Realis	Irrealis	Realis	Irrealis	Realis	Irrealis
1sg	ni-	ne-	ni-	ka-	ne/i-	na-
2sg	a-	u-	nu-	ku-	go/i-	go-
3sg	i-	i-	i-	ki-	i-	i-
1in:du	la(i)-	la(i)-	la(i)-	sa-	de-	ja-
1ex:du	ba(i)-	ba(i)-	(nu)ba-	ba(i)-	ba(i)-	ba(i)-
2du	ba(i)-	ba(i)-	(nu)ba-	ba(i)-	ba(i)-	ba(i)-
3du	la(i)-	la(i)-	la(i)-	sa-	de-	ja-
1in:pl	li-	le-	le(pe)-	kape-	le/i-, git-	la-
1ex:pl	pi-	pe-	nupe-	pe-	te/i-	tu-
2pl	pi-	pe-	nupe-	pe-	te/i-	tu-
3pl	li-	le-	le(pe)-	se(pe)-	le/i-	la-

Beyond their general family resemblance, the subject prefixes shown in Table 8 show a certain amount of formal variety, see especially the 2sg, or the plural forms. Yet once again, this formal diversity goes along with a strong structural isomorphism. In particular, all languages display the same two morphological mergers on non-singular pronouns: merger of 1<sup>st</sup> inclusive and 3<sup>rd</sup> person (with a couple of exceptions) on the one hand, and merger of 1<sup>st</sup> exclusive and 2<sup>nd</sup> person on the other. This morphological pattern is specific to the three Vanikoro languages, and is not found in neighbouring Utupua (Tryon 1994:631). This may therefore constitute an important shared innovation diagnostic of a Vanikoro subgroup (see §1.3.2).<sup>14</sup>

Finally, our three languages add to the set of free pronouns another personal category, namely 3<sup>rd</sup> PLURAL INDEFINITE. This category has a special form as a free pronoun — TEA *idi* ~ LVN *nili* ~ TNM *deli*, which may be glossed ‘people’ (cf. French *on*). When this free pronoun is the subject, the agreement marker on the verb will be an ordinary 3<sup>rd</sup> plural prefix.

### 3.1.3 Possessive classes

The three languages also agree in the morphosyntax of possession. They all encode inalienable possession identically, by juxtaposing the possessed noun and its possessor. In the absence of possessive suffixes, inalienable possessors are encoded with the

<sup>13</sup> Instead of being coded by dedicated suffixes as in POc, objects and inalienable possessors are expressed by independent pronouns, which form distinct phonological words. The only exception to this principle is the Lovono suffix *-yo* for 2sg objects and possessors, which is distinct from the free pronoun *ago*, see Table 13 below.

<sup>14</sup> Pronoun systems often provide crucial diagnostic evidence in subgrouping research (Ross 2005).

independent personal pronoun: e.g. TEA *awa ini* ~ LVN *warene yani* ~ TNM *vasare nini* /throat 3sg:INDEP/ ‘his throat’. Semantically, inalienable possession covers most body parts, plus a handful of intimate belongings, e.g. TEA *bete ene* ~ LVN *bele yane* ~ TNM *be nana* /mat 1sg:INDEP/ ‘my bedmat’.

A possessive classifier is required for alienable types of possession. Four possessive categories can be distinguished (I indicate in square brackets the Teanu form of the classifier for 1sg possessor): FOOD [*enaka*]; DRINK [*me ene*]; KINSHIP [*one*]; GENERAL possession [*enone*], used as a default.

The category of FOOD possession is larger than its label suggests. First, it covers food (including items only chewed, like areca nut and betel leaf) as well as drink, thus overlapping with the dedicated DRINK classifier. Second, it is required for most tools ('knife', 'adze', 'spear', 'hook', 'box'...), plus the generic term 'belongings'. Third, it is used for 'language' and 'custom'.

- |     |     |               |                 |              |                  |   |                  |
|-----|-----|---------------|-----------------|--------------|------------------|---|------------------|
| (7) | TEA | <i>U-labu</i> | <i>yatene</i>   | <i>enaka</i> | <i>u-lui</i>     | <i>ne mwoe</i>  | <i>enone</i> .   |
|     | LVN | <i>Ku-lo</i>  | <i>vesemele</i> | <i>aya</i>   | <i>ku-lawoi</i>  | <i>ne moe</i>   | <i>iaya</i> .    |
|     | TNM | <i>Go-nou</i> | <i>vamora</i>   | <i>ae</i>    | <i>go-lao</i>    | <i>ini nalama</i>   | <i>ie</i> .      |
|     |     | 2SG:IR-hold   | thing           | FOOD:1SG     | 2SG:IR-take.away | in house  | GEN:1SG          |
|     |     |               |                 |              |                  | 'Get my[FOOD] belongings and take them to my[GENERAL] house.' |                  |
| (8) | TEA | <i>Dapa</i>   | <i>iakapa</i>   | <i>kape</i>  | <i>le-mui</i>    | <i>piene</i>  | <i>akapa</i> .   |
|     | LVN | <i>Detu</i>   | <i>iegitore</i> | <i>gape</i>  | <i>se-moi</i>    | <i>mwamwane</i>   | <i>agitore</i> . |
|     | TNM | <i>Dato</i>   | <i>egeto</i>    | <i>mota</i>  | <i>la-muo</i>    | <i>puiene</i>   | <i>ageto</i> .   |
|     |     | PL:HUM        | GEN:1INC:PL     | FUT          | 3PL:IR-not.know  | speech  | FOOD:1INC:PL     |
|     |     |               |                 |              |                  | 'Our[GENERAL] people are going to forget our[FOOD] language.' |                  |

Once again, the three languages of Vanikoro agree perfectly on the semantic content of their formal categories.

### 3.1.4 Space directionals

Another domain where Teanu, Lovono and Tanema share identical structures is the system of space reference. Even though their adverbial directionals show impressive formal diversity (Table 9), their functional properties are parallel: they all resort to the ‘in’–‘out’ contrast to encode the sea–land axis, and use the ‘up’–‘down’ pair to encode a fixed cardinal axis, oriented towards southeast. While this system is attested elsewhere among Oceanic languages, it is distinct from the one reconstructed for POc (François 2004).

**Table 9:** Space directionals in their local and geocentric uses

Local use	Geocentric use	Teanu	Lovono	Tanema
‘in’	‘inland’	<i>takoie</i>	<i>ηate</i>	<i>eto</i>
‘out’	‘seaward’	<i>tetake</i>	<i>mwaroa</i>	<i>emo</i>
‘up’	‘toward SE’	<i>tev’ iu</i>	<i>ηau</i>	<i>iu</i>
‘down’	‘toward NW’	<i>tev’ tawo</i>	<i>lenu</i>	<i>ese</i>

### 3.2 Lexicon

The structural isomorphism so characteristic of Vanikoro languages relates not only to the morphosyntax, but also to the semantic organisation of the lexicon. I will mention successively two types of subdomain where this parallelism can be observed: the lexicon proper, and the phraseology.

#### 3.2.1 Lexical semantics

We have seen that Teanu, Lovono and Tanema provide each grammatical category (possessive classifiers, space directionals ...) with essentially the same semantic outline. The same can be said of lexical items and their meaning: when two quite distinct meanings are ‘colexified’ in one language — i.e. are expressed by the same lexical form (François 2008b) — the same pattern of colexification will almost certainly be found in the two other languages. Setting aside cases of polysemy which are shared by all or most Oceanic languages (e.g. ‘hear’–‘feel’ ...), some of the most distinctive examples of colexification are shown in Table 10.

**Table 10:** Most colexification patterns are shared across Vanikoro languages

Sense 1	Sense 2	Teanu	Lovono	Tanema
‘one, single’	‘the same’	<i>iune</i>	<i>tilu ~ tilioko</i>	<i>omwano</i>
Indefinite SG	‘another’	<i>iote</i>	<i>leka</i>	<i>keo</i>
Indefinite PL	‘others’	<i>kula</i>	<i>kule</i>	<i>kule</i>
‘all’	‘many’	<i>abia</i>	<i>maraya</i>	<i>abia</i>
‘bird’	‘friend’	<i>menuko</i>	<i>menuka</i>	<i>manuke</i>
‘light (adj.)’	‘dry’	<i>mimione</i>	<i>mimiane</i>	<i>mamiene</i>
‘black’	‘unripe’	<i>boro</i>	<i>bware</i>	<i>betika</i>

In the domain of compounding, a special case can be made regarding three adjectives, represented in Table 11. Despite their formal differences, they appear to be everywhere analysable in the same way, as if the result of calquing. Most forms are synchronically transparent, and based on the noun ‘name’. As for TEA *yasune*, it can be analysed as historically a combination of (*e*)*ya* ‘name’ and *iune* ‘one, the same’.

**Table 11:** Lexical connection between noun ‘name’ and three adjectives

English	Teanu	Lovono	Tanema	Literally
‘name’	<i>ejə</i>	<i>neje</i>	<i>niŋe</i>	
‘identical’	<i>yasune</i>	<i>neje-ilu</i>	<i>niŋe-omwano</i>	‘name-one’
‘different’	<i>ejə-iote</i>	<i>neje-leka</i>	<i>niŋe-keo</i>	‘name-other’
‘various’	<i>ejə-eja</i>	<i>neje-neje</i>	<i>niŋe-niŋe</i>	‘name-name’

Likewise, each of the three languages derives its intensifier ‘very much, too much’ — see ex.(5) — from its adjective ‘bad’.<sup>15</sup> This connection is especially noteworthy as it involves distinct roots in each language (Table 12). The second element in these compound forms is obscure.

**Table 12:** Lexical connection between adjective ‘bad’ and intensifier

English	Teanu	Lovono	Tanema
‘bad’	<i>tamwaliko</i>	<i>visale</i>	<i>vae</i>
Intensifier	<i>tamwa(liko)se</i>	<i>visale-wabeu</i>	<i>vae-pamabo</i>

#### 3.2.2 Phraseology

The structural isomorphism between the three languages of Vanikoro is equally obvious from their phraseology, i.e. the routinised way in which they connect words together. I will only mention here a couple of original cases.

The three languages have an inalienable noun for ‘body’, to which they attach a variety of meanings, including ‘genuine, true’,<sup>16</sup> and ‘beautiful’, see Table 13.

**Table 13:** The polysemy of the noun ‘body’

TEA	<i>ebele eo</i>	<i>ebele piene</i>	<i>ebele kuo</i>	<i>ebel’ ini</i>
LVN	<i>nebele -yo</i>	<i>nebele mwamwane</i>	<i>nebele nave</i>	<i>nebele yani</i>
TNM	<i>nibela go</i>	<i>nibela puiene</i>	<i>nibela goia</i>	<i>nibela nini</i>
	body 2sg	body speech	body canoe	body 3sg
	‘your body’	‘true words, truth’	‘canoe hull’ ~	‘his/her/its body’ ~
			‘beautiful canoe’	‘Wonderful!’

The neck or throat evidently constitutes, in Vanikoro, the seat of emotions and feelings.<sup>17</sup> The corresponding noun is found in a variety of formulas:

- ‘I’m angry’ is literally ‘*My throat is burning*’
- ‘I’m sad’ is literally ‘*My throat is blocked*’

Our three Vanikoro languages are strictly parallel in all these formulations. This is a fact of structural isomorphism, or calquing, as the words for ‘throat’ do not appear to be cognate across languages:

<sup>15</sup> The connection is also attested in English (*I want it badly*), and closer to Vanikoro, in the Torres languages of Vanuatu: e.g. Lo-Toga *na luwō hia*, lit. ‘it’s big bad’ = ‘it’s too big’ (François, pers. data).

<sup>16</sup> Interestingly, the languages of northern Vanuatu share the same colexification pattern between ‘body’ and ‘true’. François (2005:501) thus proposes to reconstruct, for the common ancestor of north Vanuatu languages, a protoform \*tur[i,u](y)i ‘body, trunk; the real, main, very X; really’.

<sup>17</sup> Osmond (2007) reports similar metaphors of emotions located in the larynx, for languages of the Southeast Solomons, as well as for the languages of the Trobriand Islands (after Malinowski 1922:408).

- (9) TEA *Awa* *kupa* *i-su.*  
 LVN *Warene* *gamitu* *i-tu.*  
 TNM *Vasare* *gamuto* *i-to.*  
 throat 1EX:PL:INDEP 3SG:R-blocked  
 [lit. 'Our throats are blocked.] 'We're sad ~ We're sorry.'

The sense 'like'/'want' is expressed by an unusual formula using a verb 'hit', taking the 'throat' as its subject:

- 'I like/want this' is literally '*My throat is hitting this.*'

The sentence becomes even more unusual when it is followed by an object clause ('want to do'), because it then involves a complementiser which is literally a verb meaning 'say'.<sup>18</sup> The subject of 'say' is normally the 'throat' itself (hence 3sg agreement), but occasionally it agrees syntactically with the throat's possessor:

- 'I want to [sleep]' is literally '*My throat is hitting IT SAYS I [sleep] ...'*
- or '*My throat is hitting I SAY I [sleep] ...'*
- which is often shortened to '*My throat I SAY I [sleep] ...'*'

- (10) TEA *Awa* *ene* (*i-viaene*) *ni-ko* *ne-mokoiu.*  
 LVN *Warene* *yane* (*i-piaine*) *ni-pu* *ka-mepeu.*  
 TNM *Vasare* *nana* (*i-vini*) *ni-po* *na-matou.*  
 throat 1SG:INDEP 3SG:R-hit 1SG:R-say 1SG:IR-sleep  
 'I want to sleep.'

In this case just as in all other contexts, the three languages can be translated literally, morpheme-by-morpheme, with no loss in idiomacity or change in meaning. All one has to do is keep the structural — grammatical and lexical — boxes, and swap their phonological contents.

#### 4 Addressing the paradox

In sum, the three languages of Vanikoro can be characterised by two contradictory properties. On the one hand, their fundamental genetic relatedness is blurred by a high degree of dissimilarity in the phonological forms of words, whether in the lexicon or in the morphology. But on the other hand, their grammatical categories and semantic structures show no equivalent to this formal diversity: instead, the three languages reveal perfect isomorphism, in each and every corner of their system. To paraphrase a formula by Sasse (see fn.19 below), they could ultimately be described as '*a single language with different vocabularies*'.

The question arises of what historical scenario would best explain this paradox, where divergence goes along with convergence. A simple explanation that comes to mind when accounting for the lexical diversification of cognate languages, might focus on the physical

separation between language communities. The absence, still today, of any land path relating villages across Vanikoro island, and the stories of ongoing fierce territorial fights between its three tribes, would then be understood as genuine evidence for geographical or social isolation, and thus as a possible key for the high degree of formal divergence between Teanu, Lovono and Tanema. However, several facts seem to contradict this diagnostic. First, the relatively small size of the island is at odds with the notion of a neat separation between the three tribes. And more crucially, their extreme degree of structural isomorphism is likely to reflect not only cases of shared retentions from a common ancestor, but also later linguistic convergence induced by language contact. In other words, the explanation resorting to the mere physical separation between communities does not tell the whole story.

The solution to the puzzle will probably have to be found not in the factual features of geography, but in the more subtle dimension of sociolinguistic behaviour. Indeed, a conspicuous characteristic of cultures in certain parts of Melanesia — in comparison, for example, with the Polynesian world (see Pawley 1981) — seems to be a social preference for small-scale social communities with no marked hierarchy between them, as well as a strong emphasis put on whichever anthropological or linguistic features may differ from one community to the other. Heterogeneity between villages or village groups tends to be socially valued as a way to construct a world of diversity, where each community is endowed with its own identity. In this framework, a local innovation in cultural and linguistic forms will tend to be perceived, and eventually retained, as emblematic of a specific group. Over time, this behaviour favours the emergence of cultural and linguistic divergence between erstwhile homogeneous communities. Interestingly, some language groups can be said to have only gone down this track to the point when the languages began to lose mutual intelligibility; but what is conspicuous in the case of Vanikoro languages, is that they seem to have pushed the process of differentiation far beyond that point, as though they were to keep diverging for ever.

In order to account for similar facts in other parts of Papua New Guinea, Thurston (1989), and later Ross (1996; 2001:155), have used the term ESOTEROGENY:

*Esoterogeny* is a process that adds structural complexity to a language and makes it more efficient as a medium of communication among people of the same social group, while making it more difficult for outsiders to learn to speak well. (Thurston 1989)

*Esoterogeny* arises through a group's desire for exclusiveness. (Ross 1996:184)

If the members of a community have few ties with other communities and their emblematic lect is not usually known to outsiders, then they may use it as an 'in-group' code, an 'esoteric' lect from which outsiders are consciously excluded. Innovations leading to increased complexity and to differences from neighbouring lects will be favoured. (Ross 1997:239)

One could probably discuss the degree to which such sociolinguistic processes are 'conscious', and also how they interfere with motivations of various kinds (semantic, structural, pragmatic) in bringing about change. This being said, one can probably accept the general idea behind Thurston's concept, that language differentiation in Melanesia, far from being just an accident of geographical isolation, is largely influenced by a certain social attitude whereby each group tends to produce — whether consciously or not — its own distinctive speech tradition.

<sup>18</sup> The grammaticalisation of a verb of saying into a complementiser is typologically common (Heine and Kuteva 2002; Chappell 2008). To take an Oceanic example, the verb 'say' in Araki, Vanuatu (François 2002), has exactly the same properties as in Vanikoro languages, including the persistence of a fully verbal morphology even when used as a complementiser.

Now, while this hypothesis may help explain the high amount of lexical innovation and formal divergence that took place between Vanikoro languages, it seems at odds with the remarkable stability that we've observed among their structures. I would suggest this mismatch can be explained by the different nature of the linguistic components involved here. For one thing, the phonological form of the words (Saussure's 'signifiant', Grace's 'lexification'), whether lexical or grammatical, is the component most salient and conspicuous to the speakers' conscience, and therefore most likely to be preempted by motivations based on social emblematicity. Conversely, the structural and semantic dimension of language (Saussure's 'signifié', Grace's 'content form') would fall out of reach of the speakers' immediate linguistic awareness, in a way that would make it exempt of the sociolinguistic force of *esoterogeneity*. Instead, structures tend to obey a totally contrary force, typical of language-contact situations,<sup>19</sup> that leads them to diffuse and converge: this is when multilingual speakers feel the pressure 'towards word-for-word translatable codes' (Gumperz 1971:270). The structural isomorphism that can be observed today among Vanikoro languages has the considerable advantage, for the bilingual speaker, of reducing any translation loss, thereby increasing the efficiency of cross-linguistic communication, and facilitating the cognitive processing of speech.

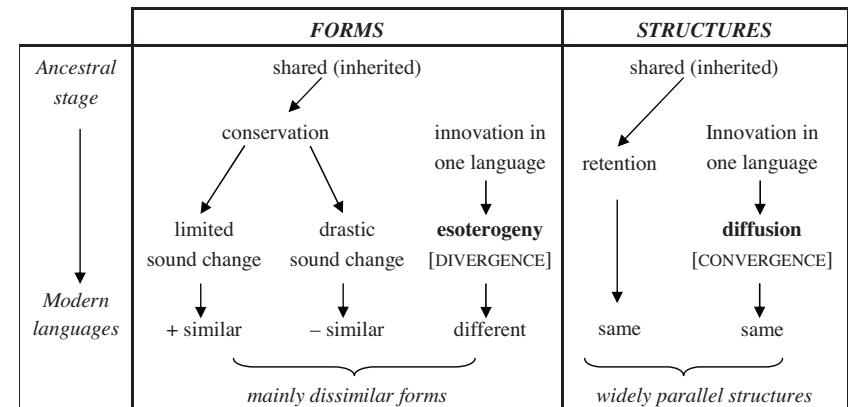
An important corollary of this whole reasoning is the necessity to distinguish two different components of language, because their evolution through history can follow quite distinct paths:

The two components of language — content form and lexification — (...) evolve independently because (...) they are responding to different selective pressures, and those selective pressures are different because the functions of the two components are different. (...) It is the lexification on which the emblematic burden ultimately falls. (Grace 1981:30)

Vanikoro illustrates an extreme case of this possible mismatch between the two components of language.

Table 14 summarises a possible scenario about the sequence of developments in the evolution of the Vanikoro languages.

**Table 14:** Different forms, shared structures among Vanikoro languages: a historical scenario



## 5 Conclusion

The comparison of Teanu, Lovono and Tanema reveals the intricacies of the island's local history. The strong isomorphism found between the structures of these languages betrays their remote common ancestry, as much as it points to a history of intense language contact which the three tribes, *nolens volens*, have lived through over the centuries. On the other hand, the actual word forms found in their vocabularies and morphology have tended to follow a powerful tendency towards diversification, in accordance with the speakers' tacit perceptions that the three communities, often caught in conflict and territorial hostilities, should sound and feel to be distinct social groups.

Overall, the paradox observed among the three modern languages of Vanikoro — *dissimilar forms, similar structures* — results from the interplay between these two contradictory forces: a socially driven push to increase language differences versus a functionally grounded tendency to minimise them.

## References

- Camden, William. 1979. Parallels in structure and lexicon and syntax between New Hebrides Bislama and the South Santo language spoken at Tangoa. In *Papers in Pidgin and Creole linguistics*, No.2, 51–117. Canberra: Pacific Linguistics.
- Chappell, Hilary. 2008. Variation in the grammaticalization of complementizers from *verba dicendi* in Sinitic languages. *Linguistic Typology* 12.45–98.
- Dillon, Peter. 1830. *Voyage aux îles de la mer du Sud, en 1827 et 1828, et Relation de la découverte du sort de La Pérouse. Par le Capitaine Peter Dillon*. Paris: Pillet Aîné.
- Dumont d'Urville, Jules Sébastien César. 1830–1834. *Voyage de la corvette l'Astrolabe, exécuté par ordre du Roi, pendant les années 1826–1827–1828–1829, sous le commandement de M.J. Dumont d'Urville, Capitaine de vaisseau*. 23 volumes. Paris: J. Tastu.

<sup>19</sup> Among many other references, see in particular Malcolm Ross' (1996, 1997, 2001) concept of *metatypy*, i.e. the typological alignment of one language to the structures of a neighbouring language, through linguistic contact. Ross (2001:149) also cites this statement by Sasse (1985): 'With advanced language contact, there arises the tendency to develop a single language with different vocabularies.'

- Dyen, Isidore, Joseph B. Kruskal, and Paul Black. 1992. *An Indo-european classification: a lexicostatistical experiment*. Transactions of the American Philosophical Society, vol. 82(5). Philadelphia: The American Philosophical Society.
- Enfield, Nicholas. 2001. On genetic and areal linguistics in mainland Southeast Asia: parallel polyfunctionality of ‘acquire’. In Alexandra Aikhenvald and R.M.W. Dixon, eds *Areal diffusion and genetic inheritance: problems in comparative linguistics*, 255–290. Oxford: Oxford University Press.
- Firth, Raymond. 1985. *Tikopia–English dictionary – Taranga fakatikopia ma taranga fakainglisi*. Auckland: Auckland University Press.
- François, Alexandre. 2002. *Araki. A disappearing language of Vanuatu*. Canberra: Pacific Linguistics.
- — — 2004. Reconstructing the geocentric system of Proto Oceanic. *Oceanic Linguistics* 43(1).1–32.
- — — 2005. Unraveling the history of the vowels of seventeen northern Vanuatu languages. *Oceanic Linguistics* 44(2).443–504.
- — — 2006. Are Vanikoro languages really Austronesian? Paper presented at the Second conference on Austronesian Languages and Linguistics. Oxford, UK.
- — — 2007. Parallel meanings, divergent forms in the North Vanuatu Sprachbund: diffusion or genetic inheritance? Paper presented at 7<sup>th</sup> International Congress of the Association of Linguistic Typology. Paris, France.
- — — 2008a. Mystère des langues, magie des légendes. In Association Salomon, ed. *Le mystère Lapérouse ou le rêve inachevé d’un roi*, 230–233. Paris: de Conti – Musée national de la Marine.
- — — 2008b. Semantic maps and the typology of colexification: intertwining polysemous networks across languages. In M. Vanhove, ed. *From polysemy to semantic change: towards a typology of lexical semantic associations*, 163–215. Amsterdam/Philadelphia: John Benjamins.
- — — forthcoming. Des valeurs en héritage: les isomorphismes sémantiques et la reconstruction des langues. In I. Choi-Jonin, M. Duval and O. Soutet, eds *Typologie et comparatisme. Hommages offerts à Alain Lemaréchal*. Orbis-Supplementa 29. Louvain: Peeters.
- — — in preparation. Language ecology and change in north Vanuatu: a tale of divergence and convergence.
- Gaimard, Joseph Paul. 1833. Extraits du Journal de M. Gaimard. In J. Dumont d’Urville, *Voyage de la corvette l’Astrolabe, exécuté par ordre du Roi, pendant les années 1826–1827–1828–1829, sous le commandement de M.J. Dumont d’Urville, Capitaine de vaisseau – Histoire du voyage*, vol. 5, 322–353. Paris: J. Tastu, éditeur.
- — — 1834. Vocabulaires des idiomes des habitans de Vanikoro. In J. Dumont d’Urville, *Voyage de découvertes de l’Astrolabe, exécuté par ordre du Roi, pendant les années 1826–1827–1828–1829, sous le commandement de M.J. Dumont d’Urville, Capitaine de vaisseau – Philologie*, vol. 1, 165–174. Paris: Ministère de la Marine.
- Grace, George. 1981. *An essay on language*. Columbia, SC: Hornbeam Press.
- — — 1990. The ‘aberrant’ (vs. ‘exemplary’) Melanesian languages. In Philip Baldi, ed. *Linguistic change and reconstruction methodology*, vol. 45, 155–173. Berlin: Mouton de Gruyter.
- Gumperz, John. 1971. Convergence and creolization: a case from the Indo-Aryan/Dravidian border in India. In *Language in social groups: essays by John J. Gumperz*, 251–273. Stanford: Stanford University Press.
- Heine, Bernd and Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Hjelmslev, Louis. 1961 [1943]. *Prolegomena to a theory of language*. Madison: University of Wisconsin Press.
- Ivens, Walter G. 1918. *Dictionary and grammar of the language of Sa'a and Ulawa, Solomon Islands*. Washington: Carnegie Institution of Washington.
- Keesing, Roger M. 1988. *Melanesian Pidgin and the Oceanic substrate*. Stanford: Stanford University Press.
- — — 1991. Substrates, calquing and grammaticalization in Melanesian Pidgin. In E.C. Traugott and B. Heine, eds *Approaches to grammaticalization*, vol. 1, 315–342. Amsterdam/Philadelphia: John Benjamins.
- Lynch, John. 2001. *The linguistic history of Southern Vanuatu*. Canberra: Pacific Linguistics.
- Malinowski, Bronislaw, 1922. *Argonauts of the western Pacific*. London: Routledge and Kegan Paul.
- Osmond, Meredith. 2007. Describing emotions and enduring dispositions in Oceanic languages. Paper presented at 7<sup>th</sup> International Conference on Oceanic Linguistics. Nouméa, New Caledonia.
- Pawley, Andrew. 1981. Melanesian diversity and Polynesian homogeneity: a unified explanation for language. In J. Hollyman and A. Pawley, eds *Essays in Pacific languages and cultures in honour of Bruce Biggs*, 269–310. Auckland: Linguistic Society of New Zealand.
- — — 2006. Explaining the aberrant Austronesian languages of southeast Melanesia: 150 years of debate. *Journal of the Polynesian Society* 115.213–256.
- — — 2007. Was there early Lapita settlement of the Solomon Islands? Bringing linguistic evidence to an archaeological debate. Paper presented at 7<sup>th</sup> International Conference on Oceanic Linguistics. Nouméa, New Caledonia.
- Ross, Malcolm D. 1988. *Proto Oceanic and the Austronesian languages of western Melanesia*. Canberra: Pacific Linguistics.
- — — 1996. Contact-induced change and the comparative method: cases from Papua New Guinea. In Mark Durie and Malcolm Ross, eds *The comparative method reviewed: regularity and irregularity in language change*, 180–217. Oxford: Oxford University Press.
- — — 1997. Social networks and kinds of speech-community event. In Roger Blench and Matthew Spriggs, eds *Archaeology and language 1: theoretical and methodological orientations*, 209–261. London: Routledge.

- — 2001. Contact-induced change in Oceanic languages in north-west Melanesia. In Alexandra Aikhenvald and R.M.W. Dixon, eds *Areal diffusion and genetic inheritance: problems in comparative linguistics*, 134–166. Oxford: Oxford University Press.
- — 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. In A. Pawley, R. Attenborough, R. Hide and J. Golson, eds *Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*, 15–66. Canberra: Pacific Linguistics.
- Ross, Malcolm D. and Åshild Næss. 2007. An Oceanic origin for Äiwoo, the language of the Reef Islands? *Oceanic Linguistics* 46.456–498.
- Sasse, Hans-Jürgen. 1985. Sprachkontakt und Sprachwandel: Die Gräzisierung der albanischen Mundarten Griechenlands. *Papiere zur Linguistik* 32.37–95.
- Siegel, Jeff. 2008. *The emergence of pidgin and creole languages*. Oxford: Oxford University Press.
- Thurston, William. 1989. How exoteric languages build a lexicon: esoterogeny in western New Britain. In Ray Harlow and Robin Hooper, eds *VICAL 1: Oceanic languages. Papers from the Fifth International Conference on Austronesian Linguistics, part 2*, 555–579. Auckland: Linguistic Society of New Zealand.
- — 1994. Renovation and innovation in the languages of north-western New Britain. In T. Dutton and D. Tryon, eds *Language contact and change in the Austronesian world*, 573–609. Berlin: Mouton de Gruyter.
- Tryon, Darrell. 1994. Language contact and contact-induced language change in the eastern outer islands, Solomon Islands. In T. Dutton and D. Tryon, eds *Language contact and change in the Austronesian world*, 611–648. Berlin: Mouton de Gruyter.
- — 2002. Buma. In J. Lynch, M. Ross and T. Crowley, eds *The Oceanic languages*, 573–586. London: Curzon.
- Tryon, Darrell and Brian D. Hackman. 1983. *Solomon Islands languages: an internal classification*. Canberra: Pacific Linguistics.
- Tryon, Darrell, ed. 1995. *Comparative Austronesian dictionary. An introduction to Austronesian studies*. Berlin/New York: Mouton de Gruyter.
- Tua, Benjamin and Peter Lincoln. 1979. *Our custom way Vanikoro: the custom book*. Auckland: University of Auckland.
- Wurm, Stephen A. 1976. The Reef Islands–Santa Cruz family. In S.A. Wurm, ed. *New Guinea area languages and language study*, vol. 2: *Austronesian languages*, 637–674. Canberra: Pacific Linguistics.



# 14 *Verbal aspect and personal pronouns: the history of aorist markers in north Vanuatu*

---

ALEXANDRE FRANÇOIS

## 1 Introduction<sup>1</sup>

The subject and object clitics which are reconstructed for Proto Malayo-Polynesian (Blust 1977) and for Proto Oceanic (Lynch, Ross and Crowley 2002:67) have been replaced in Mwotlap, as in many other languages of north Vanuatu, with a unique set of free pronouns: *nɔ* ‘1sg’; *nɪk* ‘2sg’; *kɪ* ‘3sg’... These may be used both in subject and object positions:<sup>2</sup>

- (1)      *nɔ* m-etsas *kɪ*, *kɪ* m-etsas *nɔ*.  
1sg PRF-see 3sg 3sg PRF-see 1sg  
'I saw her and she saw me.'

Most Mwotlap pronouns are morphologically invariant. However, in subject position, the 1sg pronoun shows allomorphic variation between two forms *nɔ* and *nɔk*. This uncommon alternation depends on the tense-aspect-mood (TAM) marking of the verb. Out of the twenty-five TAM categories in Mwotlap (François 2003), eight allow for free variation between the two forms, whereas in the rest of the system, they come in strict complementary distribution: ten markers require *nɔ* as their subject, while seven require *nɔk*. In fact, as we will see below, *nɔk* itself can be described as a portmanteau form indexing both person and aspect.

---

<sup>1</sup> I am grateful to COOL7 participants for their questions on an oral version of this paper (François 2007), and to Claudia Wegener and Alexis Michaud for their comments on an earlier draft.

<sup>2</sup> All transcriptions use IPA rather than local orthographies, to facilitate comparison. Note that [v] is often bilabial [β]; [j] is the palatal approximant; [ʃ] is a laminal retroflex; all voiced stops are prenasalised ([d] is [ⁿd], [gβʷʰ] is [ʷgβʷʰ], etc.). Besides abbreviations that follow the Leipzig glossing rules, AO means 'Aorist'; POT 'Potential'; PROSP 'Prospective'; STAT 'Stative'; (P)NCV '(Proto) North-Central Vanuatu'; TAM 'tense-aspect-modality'. The three-letter abbreviations for modern languages are spelled out on Map 1.

This formal variation of the 1sg pronoun depending on the predicate's TAM-marking is typologically original. It also constitutes a morphological puzzle, which I will take as the starting point for this paper. Section 2 will begin with a synchronic approach, by describing the semantic motivation of the *nɔ-nɔk* contrast in Mwotlap; this will lead to the functional definition of an aspect category labeled 'Aorist'.<sup>3</sup> Based on this definition, §3 will investigate the geographical distribution and the formal characteristics of similar aorist markers across the seventeen languages of the Banks and Torres islands. Finally, §4 will take a historical perspective, and attempt to unravel the development of aorist markers in north Vanuatu languages.

## 2 The Aorist in Mwotlap

The first question I will tackle is the functional distribution of the two allomorphs *nɔ* and *nɔk* in Mwotlap.

### 2.1 A special pronoun for the Aorist

The word order of constituents in Mwotlap is as follows:

Subject NP – (TAM clitic/prefix) – Predicate – (TAM postclitic) – (Object NP)

As far as the 1sg pronoun is concerned, its unmarked, default form is clearly the shorter allomorph *nɔ*. It is the only one found in non-subject positions—see (1)—as well as for the subject of non-TAM predicates (e.g. *nɔ na-vat�ɔ* 'I'm a teacher', *nɔ itɔk* 'I'm fine'). As for tense-marked predicates, *nɔ* combines with realis (Stative, Perfect, Comitative ...) as well as irrealis markers (Future, Potential, Counterfactual ...):

- (2) *nɔ mi-wil nu-suk.*  
1sg PRF-buy ART-sugar  
'I've bought some sugar.'

- (2') *nɔ ti-wil vih nu-suk.*  
1sg POT<sub>1</sub>-buy POT<sub>2</sub> ART-sugar  
'I can buy some sugar.'

The seven TAM categories requiring the marked form *nɔk* are the Aorist proper, the Permanisive, the Prioritive, the two Presentatives (static and kinetic), the Polite Imperative and the Prospective. Despite their semantic differences, the latter six categories are related, as they are all formally derived from the Aorist, through combination with some secondary morpheme. In other words, the marked form *nɔk* is required whenever the 1sg pronoun is the subject of a tense-marked predicate belonging to the domain of the Aorist, in the wide sense of the term. Therefore I will hereafter gloss it '1sg:AO'.

Crucially, when the predicate is an Aorist strictly speaking—as opposed to one of its derivatives—the pronoun *nɔk* is in fact the only formal TAM marking in the clause. This confirms its interpretation as a portmanteau morpheme, combining person and TAM marking:

- (3) *nɔk ñjm̩wɔl*  
1sg:AO return  
'Let me go back!' (...)

While the first person encodes the Aorist through variation of the pronoun, the 3sg does this with a prefix *ni-* on the verb, in the slot usually devoted to other aspect markers—compare (3') with (2).

- (3') *ki ni-ñjm̩wɔl*  
3sg AO:3sg-return  
'Let him go back!' (...)

Finally, all persons other than 1sg and 3sg encode the Aorist with a zero:

- (3'') *yin (Ø-)ñjm̩wɔl*  
1inc:pl (AO-)return  
'Let's go back!' (...)

The Aorist and its derivatives are the only TAM categories of Mwotlap whose marking depends on the person.

### 2.2 The semantics of the Aorist

#### 2.2.1 The various uses of the Aorist

Like several other TAM categories of Mwotlap, the Aorist is only compatible with semantically dynamic events. Its combination with a stative predicate—whether a stative verb, an adjective or a noun—forces a dynamic interpretation [see (9) and (13) below].

But the semantic information which the Aorist gives about that dynamic event is widely polysemous (François 2003:165–199). For one thing, the Aorist encodes events that come in sequence. This applies equally in past or future contexts:

- (4) *nɔk hajvey l-ñjm̩w nɔnɔn ε ki ni-etsas nɔ.*  
1sg:AO enter in-house his TOPIC 3sg AO:3sg-see 1sg  
[past context] 'I came into his house and (then) he saw me.'  
[future context] 'I'll come into his house and (then) he'll see me.'

Crucially, the same sentence (4) may be translated in English either as past or as future. This shows that the Aorist is not a tense, but an aspect, which may attach either to a realis or to an irrealis situation. In itself, (4) says nothing more than '(Let there be) my coming into his house, *and then* him seeing me ...'. What is relevant here is the relation of sequence or implication between the two successive events, regardless of how they happen to relate to the speech coordinates. The Aorist is commonly found in narratives, for any chain of events:

- (5) *ki ni-jem hay le-vet tu ki ni-kp̩wisdī how tu ni-mat.*  
3sg AO:3sg-climb up on-stone then 3sg AO:3sg-fall down then AO:3sg-die  
'He climbed up the rock, then he fell down and died.'

This use as a sequential marker in a string of events is ubiquitous in Mwotlap. Yet this form is also required in many other contexts which cannot be reduced to this explanation.

The Aorist is used for generic sentences, such as definitions or procedure descriptions—that is, utterances referring to a timeless event that bears no connection with any specific situation:

<sup>3</sup> About the term 'aorist', see the end of §2.2.2. Note that I use capitalisation, following Comrie (1976:10), whenever a given term is to be understood as a labeling convention for a morphosyntactic category specific of a given language, rather than a typological concept.

- (6) “ne-ŋmʷjajaj” ε, nik etet hejlo van ai.  
 STAT-transparent TOPIC 2sg AO:look-IPFV through thither in.it  
 ‘Transparent (means that) you see through it.’

- (7) na-mte ni-joj, na-takpʷŋmʷi ni-jejej,  
 ART-your.eyes AO:3sg-sink ART-your.body AO:3sg-shiver  
 nik mat ŋmʷol ...  
 2sg AO:die return  
 ‘[with malaria] your eyes sink, your body shivers, you faint ...’

Another example where Aorists point to virtual events whose time coordinates are left indefinite, is their use in conditional clauses (note that (8) is identical to (4) above).

- (8) nɔk hajvey l-ŋmʷ nɔnɔn ε ki ni-etsas nɔ.  
 1sg:AO enter in-house his TOPIC 3sg AO:3sg-see 1sg  
 ‘(Suppose) I came into his house (then) he would see me!’

In many cases however, the Aorist does relate to a specific situation, which may be the moment of utterance. This happens, for instance, when it represents an event as imminent:

- (9) mahi ni-kpʷŋŋ iŋm.  
 place AO:3sg-night now  
 ‘Night is about to fall.’

The imminence of the event is sometimes factual, as in (9), but quite often it is the speaker’s own projection. The Aorist thus takes on modal values, and encodes intent, optative, instructions or commands:<sup>4</sup>

- (10) nik yen mej nɔk, nɔk yen mej yin.  
 2sg AO:eat the.one there 1sg:AO eat the.one there  
 ‘You eat this one, I’ll eat that one.’

- (11) ki ni-ŋmʷol le-pnu nɔnɔn.  
 3sg AO:3sg-return in-village his  
 ‘Let him return to his village!’ [or: ‘he returned ...’, see (5)]

Strictly speaking, the Aorist cannot be said to inherently entail such illocutionary forces as desiderative or imperative, because it is also used in plain declarative sentences. In other words, just as it does not by itself convey any indication of time, it is also underspecified with regard to modality: it is found in statements as much as in hypotheses, commands or optatives. Both the time coordinates and the modal value of the Aorist thus need to be inferred from prosodic clues, and from the discourse context.

This semantic underspecification with regard to time and mood explains why the Aorist (or its derivative the Prospective) is required in modality-bound subordinate clauses: e.g., clause complements of verbs of will or manipulation, as well as purposive and consecutive clauses.

- (12) nɔ ne-njus sɔ nɔk (sɔ) in ni-ti.  
 1sg STAT-want COMP 1sg:AO (PROSP) drink ART-tea  
 ‘I want to drink some tea.’ [lit. ‘I want that I drink<sub>AO</sub>...’]  
 (13) nɔ mu-muk ne-vet 1-ep tu ki ni-vej.  
 1sg PRF-put ART-stone in-fire then 3sg AO:3sg-red.hot  
 [purposive] ‘I laid the stones on the fire so that they become<sub>AO</sub> red-hot.’  
 [consecutive] ‘I laid the stones on the fire so they became<sub>AO</sub> red-hot.’

The semantic incompleteness of the Aorist thus makes it particularly compatible with certain forms of syntactic dependency, in a way reminiscent of the subjunctive of Indo-European languages.

If a dynamic verb is reduplicated, it acquires imperfective aspectual properties, including when combined with the Aorist. This means (Comrie 1976) it may take either a *habitual* reading or a *progressive* one:

- (14) nɔk jap hij tita mino.  
 1sg:AO write to mother my  
 simple verb: perfective interpretation  
 [sequential] ‘(then) I wrote to my mother.’  
 [intent/optative] ‘Let me write to my mother!’ ...  
 (14') nɔk japjap hij tita mino.  
 1sg:AO write~IPFV to mother my  
 reduplicated verb: imperfective interpretation  
 [habitual] ‘I write to my mother (every day...)’  
 [progressive] ‘I’m writing to my mother.’

The absence of reduplication in (14) gives the verb a perfective reading, which makes it compatible with the various aspectual values reviewed so far for the Aorist: sequential, intent, etc. Conversely, reduplication in Mwotlap (François 2004) has the power to convert a perfective into an imperfective, which disrupts the impact of the Aorist marking altogether. Arguably, the latter then functions as a neutral aspect marker, whose role is simply to state the imperfective process (habitual or progressive) in relation to the context.

## 2.2.2 Defining the underlying mechanism

Despite the impressive polyfunctionality of this TAM category, it is possible to identify a constant aspectual pattern behind the variety of its contextual meanings. In all cases, the Aorist consists in representing a new event considered in itself, that is, regardless of its deictic coordinates in terms of tense or modality.

Precisely because it lacks any inherent deictic reference, this indeterminate event needs to be connected to an external point of reference—its ‘anchor’—in order to receive proper pragmatic interpretation. Quite often, the anchoring situation is easy to retrieve from the context. For example, in a string of successive events, it corresponds to the end of the previous event (4, 5). In a subordinate pattern, the dependent event will hook onto the coordinates of the main clause (12, 13). In many cases, the default reference point will be the utterance situation, whether the new event that is supposed to cling to it is presented as a statement of fact (9, 14') or as the speaker’s projection (10, 11, 14). Finally, it sometimes happens that this ‘orphan’ event in search of situational anchoring finds none, and remains

<sup>4</sup> This modal function accounts for the formal links between the Aorist, and the three modal markers derived from it (Prospective; Prioritive; Imperative).

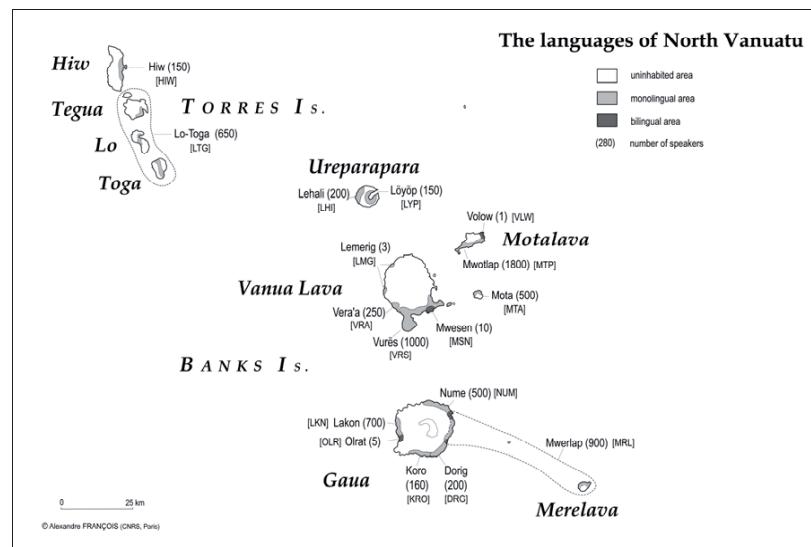
suspended in time: this is what happens when the Aorist points to a timeless event with no connection to any specific situation, as in generic statements (6, 7) or hypotheses (8).

Typologically speaking, the term ‘aorist’ has been used with various senses, and sometimes inconsistently, across language descriptions. However, the aspectological tradition that has developed, especially in France, after Benveniste (1966) and Culioni (1978), has now solidly established the notions of ‘*aoriste*’ or ‘*aoristique*’, as a verbal aspect whereby the depicted event is disconnected from the situation of utterance. Similar examples of ‘aorist’ have been described for several languages, such as Coptic (Depuydt 1993), Wolof (Robert 1996) or Berber (Galand 2003). A full typological survey of the aorist aspect still needs to be carried out.

### 3 The morphology of the Aorist in northern Vanuatu languages

Now that the semantics of the Aorist have been observed for Mwotlap on a synchronic, language-internal basis, it becomes possible to observe whether its neighbors of north Vanuatu possess a similar aspect category, and if so, how they encode it morphologically. This observation might help trace the formal history of Mwotlap’s Aorist, and especially of the unusual alternation between the two 1sg pronouns *nɔ* and *nɔk*.

Since 2003, my field investigations have precisely involved the firsthand study of all the languages of the Banks and Torres groups, of which basically nothing was known to date. Map 1 locates these seventeen languages; it indicates their current number of speakers, together with the three-letter abbreviations I propose to use for them. The remainder of this section will summarise the results of this survey with regard to the Aorist aspect.



Map 1: The languages of north Vanuatu

### 3.1 South Banks

The TAM systems observed in the five languages of Gaua, together with Mwerlap, differ significantly from that of Mwotlap. In particular, the semantic spectrum of Mwotlap’s Aorist, instead of being encompassed by a single marker, is divided in these languages into two, three or even four distinct categories, each language showing its own particular distribution (Table 1).

Table 1: Equivalents to Mwotlap’s Aorist in the six south Banks languages

	MTP ex	MRL	NUM	DRG	KRO	OLR	LKN
<i>Sequential</i>	(4, 5)	ti	tɔv	sɔ		Ø	ti
<i>Generic</i> (definitions, procedures)	(6, 7)		vɛ		v-		
<i>Subjunctive</i> (dependent, conditional)	(8, 12, 13)	sV-	te	s-		(y)a	(y)a
<i>Prospective</i> (optative, commands ...)	(9, 10, 11, 14)				s-		
<i>Imperfective</i> (habitual, progressive)	(14')		te... ti	t... ti	t... ti	ti... ti	ti... to

It would be a matter for complex discussion to decide which of these morphemes should be properly labeled ‘Aorist’, and which ones should receive a name of their own. For example, in Dorig, it is safe to call *sɔ* a Sequential, and *t... ti* an Imperfective. As for *s-*, the union of ‘generic’, ‘subjunctive’ and ‘prospective’ could be tagged Aorist, in the sense of ‘deictically indeterminate new event’, as defined above for Mwotlap. Yet it could as well, and perhaps more accurately, be called *Irrealis* or *Virtual*—a choice impossible in Mwotlap due to both the sequential and the imperfective uses.

In sum, none of these languages possess a proper aorist, in the sense defined for Mwotlap. Furthermore, all markers in Table 1 are invariable prefixes or proclitics, used for all persons. Their forms resemble neither MTP *nɔk* nor *ni-*, the origins of which will have to be sought elsewhere.

### 3.2 Central Banks

The ten remaining languages of the Banks and Torres are more promising. Indeed, each of these languages possesses a TAM category which essentially matches the Aorist of Mwotlap, encompassing all the functions of Table 1, from ‘sequential’ to ‘imperfective’;<sup>5</sup> I shall therefore use the label ‘Aorist’ everywhere. And, crucially, in each language, its formal marking depends on the person of the subject, in a way reminiscent of Mwotlap.

Let us first observe the three languages located in the central part of the Banks Islands: Mota, Mwesen and Vurès. Taking the verb ‘see’ (MTA *ilo*, MSN-VRS *il*) as an example, Table 2 illustrates the behavior of subject markers for the Aorist, in comparison with an

<sup>5</sup> In order to demonstrate this, a full set of examples should ideally be provided for each language. Unfortunately, this is impossible here due to considerations of space.

ordinary TAM marker—in this case, the Perfect.<sup>6</sup> The pattern for 1inc:pl, which is given here, exemplifies the twelve non-singular forms.

**Table 2:** Aorist inflections in three central Banks languages

Language	TAM	1sg	2sg	3sg	1inc:pl
Mota	Perfect	nau <i>me ilo</i>	ko <i>me ilo</i>	ni <i>me ilo</i>	nina <i>me ilo</i>
	Aorist	<b>na ilo</b>	<b>ka ilo</b>	<b>ni ilo</b>	nina <b>a ilo</b>
Mwesen	Perfect	na <i>me il</i>	nik <i>me il</i>	ni <i>me il</i>	nin <i>me il</i>
	Aorist	na <b>na il</b>	nik <b>a il</b>	<b>ni ni il</b>	nin <b>a il</b>
Vurës	Perfect	<b>no mi-il</b>	nik <i>mi-il</i>	ni <i>mi-il</i>	nm <i>mi-il</i>
	Aorist	<b>na il</b>	nik <b>i il</b>	<b>ni ni il</b>	nm <b>a il</b>

Taking only the non-singular forms, we would simply have an invariant clitic *a* ‘Aorist’ behaving like other TAM markers. But the singular makes the description more complex, because the marking of the Aorist differs according to the person of the subject. To be precise, two distinct cases are attested:

- The pronoun itself remains unchanged, but the Aorist clitic presents allomorphic variation according to the person of the subject. Thus for Mwesen, the Aorist is *a* for all persons, but *na* for ‘AO:1sg’ and *ni* for ‘AO:3sg’.
- The sequence {pronoun + TAM marker} found with other tenses is replaced by a single portmanteau clitic that incorporates person- and TAM-marking. Thus in Mota, *na* should be properly glossed ‘1sg:AO’, and *ka* ‘2sg:AO’.<sup>7</sup>

Vurës combines the two patterns: (a) for 2sg and 3sg, but (b) for 1sg. In fact the same complexity was found in Mwotlap, where *nøk* was to be analyzed as an aspect-indexed pronoun (‘1sg:AO’), but *ni-* as a person-indexed aspect prefix (‘AO:3sg’).

Now, MTP *ni-* is clearly the same morpheme as *ni* in these three languages. Furthermore, a connection can be drawn between that *ni* ‘AO:3sg’ and the form of the free pronoun for 3sg *ni* in Mota, Nume, Dorig and Koro. In several languages of north Vanuatu, the 3sg pronoun (*ni*, *niə...*) reflects an earlier form \*nia ‘3sg’, itself connected with *ni*. This formal connection has been blurred in Mwotlap, where the 3sg pronoun is now an innovative *k*.

These first findings thus shed light on our initial puzzle. Yet still nothing can be said about the strange form *nøk* in Mwotlap: where does this /k/ come from? The answer will appear as we continue our survey further north.

### 3.3 North Banks

Not surprisingly, a system much closer to Mwotlap can be found in Volow, an extinct dialect formerly spoken on the same island, and passively remembered by a handful of people. The structures of the two dialects are so parallel that the only differences lie in the phonological forms of the markers: to the alternation between MTP *nø* and *nøk* corresponds a

<sup>6</sup> For each language, the first row translates as ‘I have seen’ ...; the second row as ‘Let me see’, etc.

<sup>7</sup> For 3sg in Mota, the Aorist can be identified either as a Ø clitic (commuting with *a*) or as incorporated in *ni* (commuting with *ka*).

pair of forms *ne* ‘1sg’ vs *ney* ‘1sg:AO’ (see Table 3 below).<sup>8</sup> But precisely because Volow is so close to Mwotlap, it is of little help in our investigation.

More instructive findings come from the four languages of the northwest Banks area: Lehali, Löyöp, Lemerig and Vera'a. Unlike the three languages of Table 2, they do possess a trace of the velar /k/ which is found in the 1sg pronoun of Mwotlap. But, interestingly, instead of being part of the pronoun itself, the consonant /k/ is separable from it, and prefixed to the verb. This becomes obvious when the pronoun and the verb are separated by another morpheme. Compare the Prospective of Mwotlap with its form in Lehali and Vera'a:

- (15) MTP *nøk* *sɔ* mitij.  
1sg:AO PROSP sleep  
  
LHI *no* *de* *k-* mutuj.  
1sg PROSP AO:1sg- sleep  
  
VRA *no* *sɔ* *k-* mi?ir.  
1sg PROSP AO:1sg- sleep  
‘I’d like to sleep.’

This syntactic test makes it easy to define the boundary between the personal pronoun proper and the (person-conditioned) TAM-marker. Unlike Mwotlap, these northwest Banks languages thus treat the 1sg Aorist marker in the same way as their 3sg, as a prefix to the verb:

- (16) MTP *ki* *sɔ* *ni-* mtij.  
3sg PROSP AO:3sg- sleep  
  
LHI *ke* *de* *n-* mutuj.  
3sg PROSP AO:3sg- sleep  
  
VRA *di* *sɔ* *ne-* mi?ir.  
3sg PROSP AO:3sg sleep  
‘He’d like to sleep.’

Lemerig does not allow any element between the pronoun and the (inflected) verb, which makes it impossible to conduct the test illustrated by (15). For example, ‘Let me sleep’ will take the ambiguous surface form /nækmi?ir/, which could be parsed *næk mi?ir* <1sg:AO|sleep> as in Mwotlap, or *næ k-mi?ir* <1sg|AO:1sg-sleep> as in Vera'a. My Lemerig corpus shows 38 instances of a 1sg Aorist, out of which 37 show this ambiguity. Luckily, one sentence has two Aorists chained together, a context where the personal pronoun may be dropped. This single example gives the solution to the puzzle, and highlights the structural difference between Lemerig and Mwotlap:

- (17) LMG *næ* *k-oen* sur ε (*næ*) *k-mi?ir*.  
1sg AO:1sg-lie down LNK (1sg) AO:1sg-sleep  
  
MTP *nøk* ε *hi* ε (*nøk*) mitij.  
1sg:AO lie down LNK (1sg:AO) (AO:)sleep  
‘Let me lie down and sleep.’

<sup>8</sup> Note that the correspondence between MTP /k/ and VLW /ŋ/ syllable-finally is regular, and reflects a former prenasalised voiced stop [⁹g] (noted \*g).

Finally, the case of Löyöp is slightly tricky, because it is a hybrid of both patterns. On the one hand, just like its neighbours but unlike Mwotlap, Löyöp has kept a 1sg Aorist prefix of the form *k-*. This form appears when the clause lacks the free pronoun *nø*:

- (18) LYP nø nye möjøs se *k-suwji* n-kp<sup>w</sup>en.  
1sg STAT want SUB AO:1sg-cast ART-net  
'I want to go net-casting.' [lit. I want that I<sub>AO</sub> cast the net]

However, Löyöp has also taken the same path as Mwotlap, in that the combination of the free pronoun *nø* and of the *k-* prefix has been resegmented, giving rise to a new, unanalysable pronoun *nøk*, with an unpredictable vowel:

- (19) LYP nø te pil tʃe, *nøk* døn se n-jø-k.  
1sg NEG<sub>1</sub> steal NEG<sub>2</sub> 1sg:AO think SUB ART-CLPOS-1sg  
'I didn't steal it, I thought it was mine.'

Taking the verb meaning 'see' again as an example,<sup>9</sup> Table 3 shows the Aorist morphology for the six languages under comparison here—including Mwotlap and Volow—thereby covering the whole 'north Banks' area. The hybrid case of Löyöp appears in the middle.

**Table 3:** Aorist inflections in six north Banks languages

Language	1sg	2sg	3sg	1inc:pl
Lemerig	nø k-?et	næk (ø-) ?et	ti <b>n</b> -?et	yæt (ø-) ?et
Vera'a	nø k-?n	niki (ø-) ?n	di <b>n</b> e-?n	yidi k-?n
Lehali	nø k-et	nek (ø-) et	ke n-et	yen (ø-) et
Löyöp	<b>k</b> -et / <i>nøk</i> et	nij (ø-) et	kje <b>n</b> -et	jen (ø-) et
Mwotlap	nøk (ø-) et	nɪk (ø-) et	kɪ <b>n</b> i-et	yɪn (ø-) et
Volow	nøŋ (ø-) et	nɪŋ (ø-) et	gɪ <b>n</b> -et	yɪŋ (ø-) et

The problem raised by *k-* in Vera'a non-singular forms will be addressed in §4.2.

## 3.4 The Torres Islands

### 3.4.1 Two sets of personal markers

I will end this survey of Aorist markers in north Vanuatu with the two languages of the Torres group. In comparison with the Banks languages, the Aorist inflection in Hiw and Lo-Toga is morphologically richer. Not only are there specific (non-zero) morphemes for each person in the singular, but also for non-singular subjects, including different forms for the dual and for the plural (the Torres languages have lost the trial).

Table 4 lists the complete sets of personal subject markers for the two languages. On the left are given the free pronouns; on the right, the set of person-indexed Aorist clitics.

**Table 4:** Full pronouns vs Aorist clitics in the two Torres languages

	Hiw full pronoun	Aor. clitic	Lo-Toga full pronoun	Aor. clitic
1sg	nøkə	kə	nøkə <sub>(LO)</sub> / nøkə <sub>(TGA)</sub>	kə
2sg	ikə	wet ~ wek	nikə	wə
3sg	nina	nə	niə	ni
1in:du	tøgLə		tor	
1ex:du	kamagLə	{ gLə	kəmər	{ or
2du	kimigLə		kəmor	
3du	søgLə		hor	
1in:pl	titə	ta	ṭøya <sub>(LO)</sub> / yɪṭə <sub>(TGA)</sub>	
1ex:pl	kama	—	kəmə(m)	{ yə
2pl	kimi	—	kəmi	
3pl	sisə	sə	nihə	

The first obvious observation is that these two languages possess interesting clues for our study: the form of the 3sg clitic (*nə/ni*) recalls the prefixes *n-* or *ni-* we saw in the Banks languages; and the 1sg clitic *kə* is reminiscent of the prefix *k-* shown in Table 3 above. Finally, the 1sg pronoun *nøkə* strikingly resembles Mwotlap *nøk*, a point which warrants a discussion of its own (see §4.2.2).

### 3.4.2 TAM markers or light pronouns?

Just as in Banks languages, the function of the clitics of Table 4 is essentially to encode a TAM category, the Aorist. This status is proven by the comparison of (20) and (20'). Both sentences show serial verb constructions, one with the Potential, the second with the Aorist.

- (20) LTG kəmi *si* ḷʷulə *si* mətūr.  
2pl POT return POT sleep  
[Potential] 'You (pl) may go back and sleep.'  
  
(20') LTG kəmi *yø* ḷʷulə *yø* mətūr.  
2pl AO:PL return AO:PL sleep  
[Aorist] 'You (pl) go back and sleep!'

But there is further complexity. Amongst the clitics of Table 4, only two (*ni* and *yø* in Lo-Toga, none in Hiw) may be immediately preceded by a free pronoun, as in (20') *kəmi yø*. All other clitics must be deleted in presence of the free pronoun, in which case the latter is directly followed by the verb. As a result, most Aorist sentences, when they include the free pronoun, appear to be unmarked (or zero-marked) for TAM. Conversely, the clitics are restricted to those clauses that lack a free pronoun. This happens typically in a string of clauses, when the pronoun is mentioned only with the first verb [see (17)]:

- (21) LTG *nikə* (Ø) ḷʷulə *wə* mətūr.  
2sg (AO) return AO:2sg sleep  
  
HIW *ikə* (Ø) ḷʷujə *wət* mitiġL.  
2sg (AO) return AO:2sg sleep  
'You (sg) go back and sleep!'

<sup>9</sup> Once again, '1inc:pl' stands for all non-singular forms.

A superficial analysis of (21) would probably have posited only one marker for the Aorist (*zero*), and then two sets of pronouns: ‘heavy’ pronouns for the main or first clause in a string, followed by ‘light’ pronouns in secondary and other dependent clauses. In that framework, it would have made sense to label these clitics ‘secondary subject pronouns’. This interpretation is appealing, and could perhaps be proposed for Hiw; but in Lo-Toga, it seems to be contradicted by (20’). For the sake of consistency, it is thus safer to analyze *wə* in (21) not as a personal pronoun, but as a (person-indexed) aspect clitic. For most subjects,<sup>10</sup> deletion rules must be formulated, whereby the sequence {pronoun+clitic} simplifies to {pronoun}, e.g. \**nikə wə* V ⇒ *nikə* V.

Because Hiw operates this deletion rule for all its pronouns, its Aorist clitics seldom show up in fluent speech, as they are restricted to subordinate or secondary clauses; and even in that case they are optional, being often replaced by the full pronouns. The situation is very different in Lo-Toga, where the clitics are extremely productive, and massively represented in my corpus. This productivity of Lo-Toga clitics has two reasons: first, the two clitics *ni* and *yə* cannot be deleted, and are pervasive in speech; second, each clitic also appears as a constituent element in three compound TAM markers historically derived from the Aorist: Prospective ⟨te + Cl.⟩, Time Focus ⟨Cl. + akə⟩, and Future ⟨te + Cl. + akə⟩ – e.g., *nikə te w'* *akə mətūr* ‘you will sleep’.

In sum, in Hiw and Lo-Toga, one identifies a clause as Aorist either because it displays an Aorist clitic, or because it consists of {free pronoun + zero-marked verb}. For example, the Lo-Toga sentence (21) shows two Aorist predicates: *nikə η "ulə* ‘you return<sub>[AO]</sub>’ and *wə mətūr* ‘you sleep<sub>[AO]</sub>’.

#### 4 The historical perspective

The eleven languages endowed with a genuine aorist (§3.2 to §3.4) show such solid formal similarities that they obviously share a common history. In this section I will endeavor to reconstruct a set of aorist markers for their common (post-POc) ancestor language. Logically, this protolanguage should be Proto North-Central Vanuatu, the proposed ancestor for the majority of Vanuatu languages (Clark 1985). But since my reconstruction is reflected only in the northernmost languages of this family—those spoken in the Banks and Torres groups—it could well represent a subgroup within NCV, the precise limits of which would need to be confirmed by further diagnostic evidence.

##### 4.1 Reconstructing the set of Aorist proclitics

First, this pre-modern system possessed a set of free pronouns. They can be reconstructed for these northern Vanuatu languages (Clark 1985; Lynch and Ozanne-Rivierre 2001:38): \**nau* ‘1sg’, \**nigo* ‘2sg’, \**n(a)ia* ‘3sg’; \**kida* ‘1inc:pl’; \**ga(ma)mi* ‘1exc:pl’; \**gamuyu* ‘2pl’; \**n(a)ira* ‘3pl’.

The system can also be reconstructed with a set of person-indexed aorist markers, consisting of proclitics preceding the verb. Taking into account the phonological history of this area—especially the phenomenon of vowel reduction (François 2005)—the most plausible reconstructions are as follows:

<sup>10</sup> In Lo-Toga, this deletion rule applies to 1sg, 2sg, and dual forms. To this list, one must add *təyə* ‘1inc:pl’ in the Lo dialect of Lo-Toga: compare Toga *yitə yə η "ulə* with Lo *təyə (Ø) η "ulə* ‘Let’s go back’.

- 1sg: all languages point to the ‘nasal grade’ \*g [ŋ] (see fn.8). As for the vowel that followed this consonant, comparative evidence (see below) suggests it must have been /u/, hence a protoform \*gu.
- 2sg: a single vowel, probably \*u (reflected as /w-/ in the Torres; /i/ in Vurës; zero in most languages).
- 3sg: \*nV, probably \*ni (reflected as /ni/, /nə/, /ne/ or /n-/).
- non-singular: the modern forms /yə/, /a/ and Ø suggest a reconstruction \*(k)a.<sup>11</sup> The dual forms of the Torres may result from a local innovation, perhaps \*\*(k)a-ru (?).

The semantic array reconstructible for this set of aorist clitics most probably coincided with the observations made for modern Mwotlap (§2.2) and its neighboring languages. Their function was to construe a ‘deictically indeterminate new event’—a definition which encompasses the functions of *sequential*, *generic*, *subjunctive*, *prospective* and (with reduplication) *imperfective*.

Syntactically, these proclitics {\*gu, \*u, \*ni, \*(k)a} occupied the same slot as other TAM markers. They were preceded by the free pronoun in main clauses, or in the first clause of a chain (serialised verbs, narratives); yet they appeared on their own in dependent or secondary clauses (e.g., same-subject sequential clauses). It is thus possible to reconstruct sentences such as (22):

- (22)    \*nau *gu=*    mule *gu=*    maturu  
       1sg AO:1sg= return AO:1sg= sleep  
       ‘So I went back and slept.’ ~ ‘Let me go back and sleep!’ ...
- \*nigo *u=*    mule *u=*    maturu  
       2sg AO:2sg= return AO:2sg= sleep  
       ‘So you went back and slept.’ ~ ‘You go back and sleep!’ ...
- \*nia *ni=*    mule *ni=*    maturu  
       3sg AO:3sg= return AO:3sg= sleep  
       ‘So he went back and slept.’ ~ ‘Let him go back and sleep!’ ...
- \*kida *(k)a=*    mule *(k)a=*    maturu  
       1inc:pl AO:non.sg= return AO:non.sg= sleep  
       ‘So we went back and slept.’ ~ ‘Let’s go back and sleep!’ ...

As far as their origin is concerned, the singular forms {\*gu, \*u, \*ni} are reminiscent of two sets of personal markers:<sup>12</sup>

- the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> person singular forms of the possessive suffixes, whose protoforms in north Vanuatu are {\*-gu, \*-u<sup>13</sup>/-mu, \*-na}, from POc {\*-gu, \*-mu, \*-na};

<sup>11</sup> The Sungwadia language of north Maewo also has a series of aorist prefixes (Agnès Henri pers. comm.). The plural form is *ye*, which tends to confirm the consonant in our reconstruction \*(k)a.

<sup>12</sup> A formally very similar set of subject pronouns 1sg *gu*, 2sg *mu*, 3sg *ni*... is attested in modern Marovo, a Northwest Solomonese language spoken in the Solomon Islands (Evans 2008:404).

<sup>13</sup> For 2sg \*-u, see François (2005:486). This local variant of the suffix may have played some influence upon the 2sg proclitic \*u.

- one of the sets reconstructed for the POc subject proclitics, namely {\*ku, \*mu, \*(y)a/ñɑ}, which ultimately reflect PMP genitive pronouns (Blust 1977; 2003). As for the non-singular prefix \*(k)a, it is reminiscent of a subject clitic \*ka[i] '1exc:pl' reconstructed for some Oceanic interstage languages (Lynch et al. 2002:68).

Given the nature of the Aorist clitics, the subject clitics are a more likely source than the possessive suffixes. There is still some debate about the precise function of these POc proclitics. According to Kikusawa (2005), they retained in POc their earlier function as ergative subjects (i.e., 'A' in divalent clauses) as opposed to intransitive subjects ('S' in monovalent clauses). Lynch et al. (2002:68) suggest that this function was probably 'being lost when POc broke up'.

But one still has to explain how a set of ordinary subject pronouns should have evolved into TAM-marking clitics. A tentative hypothesis would suggest these subject markers once became specialised in subordinate or other dependent clauses while main declarative clauses eventually generalised the use of free pronouns. As a result, what were once genuine subject pronouns in clauses otherwise unmarked for TAM, eventually grammaticalised into subjunctive-like TAM markers. This hypothesis would account for the affinities of Aorist predicates with syntactic dependency, discourse backgroundedness, and TAM indeterminacy. But this is mainly speculation at this stage; the functional connection at stake here definitely warrants further investigation.

## 4.2 From the protosystem to modern languages

After tentatively reconstructing the protosystem of Aorist clitics in the protolanguage ancestral to the Torres and Banks groups (either PNCV or one of its branches), I will end this study with an overview of the various paths of evolution that historically led to the modern systems.

### 4.2.1 Phonological attrition and affixation

The phonological process of *unstressed vowel deletion*, which massively affected the languages of the whole Banks and Torres area (François 2005), explains why former \*CV clitics are generally reflected as a single consonant in modern languages (\*gu > /k/ or /ŋ/; \*ni > /n/...), and also why the two vowel-only clitics are so often reflected as *zero*. Lemerig shows the expected reflexes in this regard:

	Protosystem	>	Lemerig
'let me go back'	*'nau gu=¹mule	>	nœ k-mul
'you go back'	*'nigo u=¹mule	>	næk Ø-mul
'let him go back'	*('nia) ni=¹mule	>	(ti) n-mul
'let's go back'	*'kida a=¹mule	>	yæt Ø-mul

This vowel reduction process explains why the former clitics generally became prefixes. It also accounts for the formal convergence between 2sg \*u and the non-singular clitic \*a—variant of \*(k)a—in the form of *zero*. As for Mwesen and Mota (Table 2), this convergence rather results from the spread of the non-singular clitic a to 2sg (MTA *ka* <\*ko a); see also below for the case of 1sg.

Among the four clitics {\*gu, \*u, \*ni, \*(k)a}, only 3sg \*ni is preserved in the eleven languages endowed with a genuine 'aorist', from Hiw all the way down to Mota. On the other hand, \*(k)a seems to show the widest historical extension, as its reflexes are scattered in various places from Lo-Toga yø down to Lakon/Olrat (y)a (Table 1).

### 4.2.2 The intricate destiny of 1sg \*gu

As for 1sg \*gu, it has survived in the eight (or nine: see below for Vurës) northernmost languages of the area, but has left little trace in any other language further south. To the best of my knowledge, the only other NCV language with a reflex of the subject marker \*gu is Tamabo, with ku '1sg' (Jauncey 2002:610).

The detailed evolution of this form \*gu in north Vanuatu languages is intricate. The former proclitic \*gu= is regularly reflected as a proclitic kœ= in the Torres languages, or a prefix k- in northwest Banks languages (Table 3). The 1sg pronoun alternation (nɔ/nøk) of Mwotlap—the starting point of this study—results from a process of reanalysis: the sequence /nɔ + k-/ (with no intervening element), was so overwhelmingly frequent in speech, that it was eventually resegmented as nɔk; the former verbal clitic became accreted to the preceding pronoun, while the verb itself appeared in its bare form. The zero form taken by the Aorist with other persons (Table 3) probably added to the pressure towards morphological leveling. Volow followed a similar path, leading to an allomorphic variation between ne and ney; as well as Löyöp, with the forms nɔ and nøk.

So far, the accreted form nøk in Mwotlap has remained restricted to its original function—namely, 1sg subject of an Aorist predicate—without much affecting the regular form of the 1sg pronoun nɔ in other contexts. However, a slight tendency towards the expansion of nøk is perceptible with other TAM categories. Besides the six markers that are formally derived from the Aorist (§2.1), the use of nøk is increasingly frequent, albeit optional, for as many as ten TAM markers, historically unrelated to the Aorist—for example, the negative markers.

This tendency for the accreted form to gain ground over the original 1sg pronoun is only incipient in Mwotlap, but has reached its final stage in the two Torres languages. Due to their high frequency in discourse, the augmented forms—nɔkə in Hiw and Lo, nekə in Toga—have now become the ordinary 1sg pronoun regardless of the predicate's TAM-marking, and indeed for all functions (subject, object, etc.).<sup>14</sup> In other words, the pronoun resulting from the coalescence of \*nau gu has replaced \*nau in all positions.

Unlike Mwotlap, the coalescence of \*nau and \*gu in the Torres languages did not eliminate \*gu as an independent morpheme (ka) in the system. But the fact it was historically incorporated into the free pronoun explains why the two forms are incompatible (\*nɔkə kə is ungrammatical). Crucially, this is the key to many of the so-called 'deletion rules' which are required in the synchronic description of the Torres languages (§3.4.2). A similar process of accretion is the origin of the form tɔyø used in the Lo dialect, the only plural form in Lo-Toga to be incompatible with yø (see fn.10): tɔyø < \*(y)i tɔ yø < \*kida ka. Likewise, in Hiw, ninə has incorporated the clitic nə (ninə < \*ni(ə) nə < \*nia ni); titə has incorporated tə; sisə has incorporated sə; and so on. Hiw is the language where the contamination of the pronoun system with Aorist clitics has been maximal.

<sup>14</sup> The earlier forms \*nɔ and \*ne only survive vestigially, in the 1sg possessive of Lo (mi-nɔ) and Toga (mi-ne)—etymologically 'with me'.

Interestingly, the 1sg Aorist clitic \**gu* has also undergone another reanalysis of a completely different kind. In modern Vera'a, the prefix *k-* encodes the Aorist not only for 1sg, but for all non-singular persons as well (Table 3). It looks as if this prefix were in the process of being reanalyzed as the general marker for Aorist, thereby losing its original connection with the first person; ironically, this is the exact opposite of the Torres evolution. Even 2sg and 3sg seem threatened by the expansion of VRA *k-* to all persons, judging by the attestation of *nik sɔk-van* and *di sɔk-van* as alternatives to the regular *nik sɔ Ø-van* and *di sɔ ne-van* (respectively ‘you/he should go’).

Finally, this generalisation of /*k*/ to non-singular persons in Vera'a possibly provides the explanation for a non-etymological /*k*/ found in several non-singular pronouns in Vera'a's closest neighbor, Vurës: *kemek* ‘1exc.pl’; *kumurok* ‘1exc.du’; *durok* ‘1inc.du’. If this hypothesis is right, then Vurës would combine the evolution patterns of Vera'a (spreading of \**gu* to non-singular persons) and of the Torres languages (accretion of \**gu* to the preceding prounoun, and generalisation of the augmented form to all functions, regardless of the verb's aspect). If this hypothesis is true, then these three non-etymological /*k*/ are the only trace left by \**gu* in Vurës. Indeed, this language—like Mota and Mwesen—encodes its 1sg Aorist subjects with an innovative form *na* (Table 2), surely the result of the coalescence of *no* <\**nau* and *a* <\*(*k*)a.

## 5 Conclusion

Out of the seventeen languages of north Vanuatu, eleven share an aspect category labeled ‘Aorist’, whose function is to represent a new event regardless of its deictic coordinates in terms of tense or modality. Despite their differences, the various morphosyntactic patterns that encode this Aorist can be shown to derive ultimately from a single protosystem: a set of four portmanteau proclitics {\**gu*-\**u*-\**ni*-\*(*k*)a} combining aspect marking and subject agreement.

The reconstruction proposed in this paper not only helps to explain the development of these languages' TAM systems, but also proves indispensable in unravelling the historical morphology of personal pronouns in this part of Oceania. The next challenge is now to define the precise link—both formal and functional—that connects these four clitics to the set of personal pronouns which Robert Blust reconstructed for the remote ancestors of these languages.

## References

- Blust, Robert. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *University of Hawai'i Working Papers in Linguistics*, 9/2:1–15.
- . 2003. Three notes on early Austronesian morphology. *Oceanic Linguistics* 42-2:438–478.
- Benveniste, Émile. 1966 [1959]. Les relations de temps dans le verbe français. In *Problèmes de Linguistique Générale*, vol.1, 237–250. Tel. Paris: Gallimard.
- Clark, Ross. 1985. Languages of North and Central Vanuatu: Groups, chains, clusters and waves. In Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 199–236. Canberra: Pacific Linguistics,
- Comrie, Bernard. 1976. *Aspect. An introduction to the study of verbal aspects and related problems*. Cambridge: Cambridge University Press.
- Culioli, Antoine. 1978. Valeurs aspectuelles et opérations énonciatives : l'aoristique. In J. David and Robert Martin, eds *Actes du colloque sur la notion d'aspect*, 182–193. Metz: Klincksieck.
- Depuydt, Leo. 1993. *Conjunction, contiguity, contingency: on relationships between events in the Egyptian and Coptic verbal systems*. New York: Oxford University Press.
- Evans, Bethwyn. 2008. Subject agreement in Marovo: diachronic explanations of synchronic conditions. *Oceanic Linguistics* 47–42, 383–408.
- François, Alexandre. 2003. *La sémantique du prédicat en mwotlap (Vanuatu)*. Collection Linguistique de la Société de Linguistique de Paris, 84. Paris, Louvain: Peeters.
- . 2004. La réduplication en mwotlap : les paradoxes du fractionnement. In E. Zeitoun, ed. *Les langues austronésiennes. Faits de langues* n°23–24:177–194.
- . 2005. Unraveling the history of the vowels of seventeen northern Vanuatu languages. *Oceanic Linguistics* 44–2:443–504.
- . 2007. Verbal aspect and personal pronouns: the history of aorist markers in north Vanuatu. Paper presented at 7th International Congress on Oceanic Linguistics (COOL7), Université de Nouvelle-Calédonie, Nouméa, July 2007.
- Galand, Lionel. 2003. L'aoriste berbère, l'aspect et les valeurs modales. In Jérôme Lentini and Antoine Lonnet, eds *Mélanges David Cohen*. Paris: Maisonneuve and Larose, 235–246.
- Jauncey, Dorothy. 2002. Tamabo. In Lynch, Ross and Crowley, eds 2002:608–625.
- Kikusawa, Ritsuko. 2003. A new view of the Proto Oceanic pronominal system. *Oceanic Linguistics* 42-1:161–186.
- Lynch, John and Françoise Ozanne-Rivierre,. 2001. Some shared developments in pronouns in languages of Southern Oceania. *Oceanic Linguistics* 40-1:33–66.
- Lynch, John; Malcolm Ross and Terry Crowley. 2002. *The Oceanic languages*. Richmond: Curzon Press.
- Robert, Stéphane. 1996. Aspect zéro et dépendance situationnelle: L'exemple du Wolof. In C. Müller, ed. *Dépendance et intégration syntaxique (Subordination, coordination, connexion)*. Linguistische Arbeiten. Tübingen: Niemeyer, 153–161.



## **Pragmatic demotion and clause dependency**

On two atypical subordinating strategies  
in Lo-Toga and Hiw (Torres, Vanuatu)

Alexandre François

Langues et Civilisations à Tradition Orale,  
Fédération Typologie et Universaux Linguistiques, CNRS

Despite the wealth of subordinators in Hiw and Lo-Toga (Oceanic, north Vanuatu), two of their Tense-Aspect-Mood categories – the *Subjunctive* and the *Background Perfect* – can do without them, and encode clause dependency by themselves. A pragmatic hypothesis is proposed to account for this clause-linking faculty. The Subjunctive differs from other irrealis categories insofar as it lacks any specific illocutionary force; the Background Perfect labels its predicate as informationally backgrounded. In both cases, the clause lacks certain key properties (illocutionary force; informational weight) which are normally required in pragmatically well-formed utterances. This *pragmatic demotion* makes the clause dependent on external predication, which naturally results in syntactic subordination. This case study illustrates how syntax can be reshaped by the pragmatic parameters of discourse.

### **1. Two cases of subordination with no subordinator**

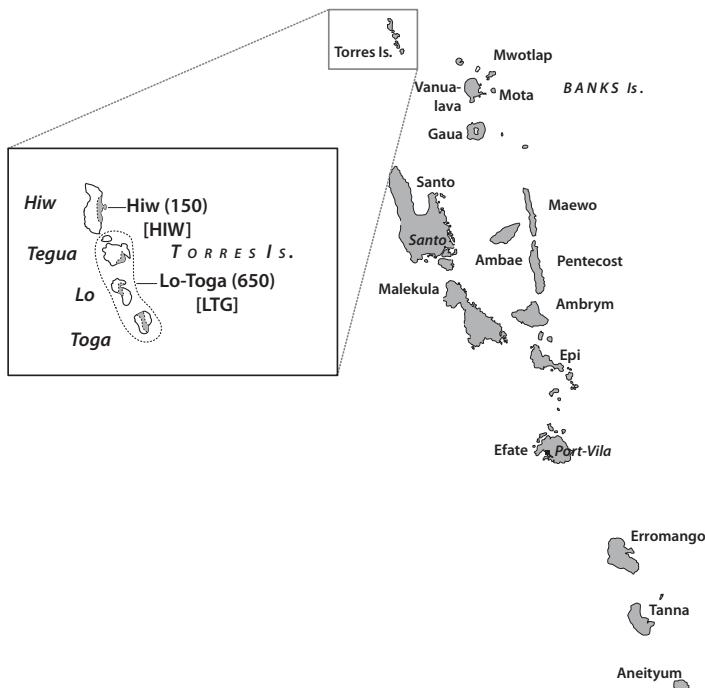
#### **1.1 The Torres languages**

The Torres islands form a small island group located at the northwestern tip of the Republic of Vanuatu (formerly New Hebrides), in the south Pacific (Map 1).<sup>1</sup> Two

---

1. The present work originates in a talk I gave in 2006 for the research group *Typology of interclausal dependencies* (Fédération Typologie et Universaux Linguistiques), led by Isabelle Bril. I am grateful to her, as well as to Alexis Michaud, Claudia Wegener and Johan van der Auwera, for their helpful comments on earlier versions of this paper. The data presented in this chapter were collected by the author during several field trips to the Torres islands, in 2004, 2006 and 2007. The financial support of the LACITO – CNRS, as well as of the French *Ministère de la Recherche* (ACI “Jeunes Chercheurs”), is also gratefully acknowledged.

Oceanic languages are spoken there: Hiw by 150 speakers, and Lo-Toga – itself consisting of two very close varieties Lo and Toga – by 650 speakers. They have never been the subject of any published grammatical description.



Map 1. The two Torres languages, at the northwestern tip of Vanuatu

Hiw and Lo-Toga differ from each other in many regards, whether in their phonology, their lexicons, or details of their grammars – enough to make them clearly distinct, mutually unintelligible languages. Nevertheless, they also share parallel structures in most domains of their morphosyntax, their phraseology, and more generally the way they categorize meaning into forms. This linguistic isomorphism between the two Torres languages is due both to their common ancestry, and to a history of sustained social and cultural contact which their communities have long had with each other. The linguistic phenomena to be discussed in the present chapter belong to those many structures which are shared by the two languages: this is why I will treat

them together here, and illustrate each phenomenon with evidence taken alternatively from Hiw and from Lo-Toga.<sup>2</sup>

While these two Torres languages also have a lot in common with the languages of the Banks group – and those of Vanuatu in general – spoken further south (Map 1), they present many specific developments, which tend to give them a grammatical profile of their own. This is especially true of the topic I will discuss here, namely the morphosyntactic strategies for encoding clause dependency and subordination. Generally speaking, as we shall see in Section 2, the various types of dependency between clauses or predicates (subordination, coordination...) are expressed – quite classically – by a variety of conjunctions and other overt morphemes that are more or less dedicated to this clause-linking function. Yet, despite the wealth of these formal devices, these two languages have also developed certain patterns of clause dependency that lack any formal subordinator.

## 1.2 Parataxis or subordination?

Considered superficially, each of the following sentences consists simply of a string of two clauses, with no formal indication whatsoever of their syntactic relationship.<sup>3</sup>

- (1) HIW Ne temêt on tō yaqe me  
ART devil SBJV go:SG appear hither  
ñwē ne, tekñwa voyi.  
like this people AOR:run.away  
[lit. The devil would appear like this, people ran away.]  
'(Whenever) the devil appeared, people would run away.'
- (2) LTG Ne gehuh ve kerkur tēle si mat mēt.  
ART coconut.crab BKPF<sub>1</sub> ITER~crunch person BKPF<sub>2</sub> CPLT die  
[lit. The coconut crab has devoured people has died.]  
'The coconut crab (which) had devoured people was dead.'

One might propose to see in these two sentences examples of simple clause parataxis (cf. Noonan 1985: 55), or perhaps of verb serialization. In fact I will show that (1) and (2) rather illustrate genuine patterns of syntactic subordination, in the full sense of the term.

While such instances of apparent clause parataxis are frequent in the spontaneous speech of the two Torres languages, they are much more constrained than they at first appear, and depend on the Tense-Aspect-Mood marking (TAM) on the verbs. Among

2. When a given fact is unique to one of the two languages, this will be stated explicitly: see for example the resultative construction, which exists only in Lo-Toga.

3. The spelling conventions adopted for the two Torres languages include the following: *g* = [ɣ]; *ñ* = [ɲ]; *ñw* = [ɲʷ]; *q* = [kʷ]; *d* = [t]; *r* = [ɾ]; *o* = [ɔ]; *ō* = [θ]; *e* = [ə]; *ē* = LTG [ɛ], Hiw [e]; *ē* = LTG [e], Hiw [i].

the many TAM categories – about sixteen – present in each of these two languages, *only two* appear to trigger seemingly paratactic structures of this sort. One belongs to the domain of irrealis modality, and is called the *Subjunctive* ('SBJV'); this appears as *on* in the Hiw sentence (1). The other belongs to the set of realis TAM markers, and more precisely to the perfect aspect; due to its particular properties, I propose to label it the *Background Perfect* ('BKPF') – expressed as *ve... si* in (2).

Ultimately, these two TAM categories – each one for distinct reasons and through different mechanisms – can be said to convey the status of their clause as being syntactically subordinate to another main clause. In other words, apparently paratactic sentences such as (1)–(2), even though they may lack any formal conjunction, can still be said to be *formally marked as subordinate*: this information is conveyed by the TAM marking on the verb, instead of being coded by clause linkers. Thus, the first clause in (1) is marked as a dependent clause by the presence of the Subjunctive; likewise, the first predicate phrase of (2) is formally identifiable as a subordinate (relative) clause through the use of the Background Perfect.

### 1.3 Formal properties, functional mechanisms

In this study, I intend not only to establish the empirical facts for these two undescribed languages, but also to propose a functional interpretation and discussion. I will adopt a functionalist perspective on this set of linguistic facts, and suggest that the syntactic effect of these two TAM categories, rather than just a purely formal property, can be shown to result from their semantic and pragmatic values.

In a nutshell, the core function of the Subjunctive in the Torres languages<sup>4</sup> is to represent a virtual state of affairs, with no further information on modality or illocutionary force. This pragmatic indeterminacy is fundamentally the reason why a subjunctive clause will need to attach itself to another clause, which provides it with the modality value it lacks. Likewise, the Background Perfect can be defined as a perfect aspect which demotes its predicate from the scope of the informational focus. Due to this backgrounded status, the predicate then needs to attach itself to another element under focus, in order to form a valid utterance.

The two cases thus appear to follow similar logic. Intrinsically, each of these two TAM markers combines its purely semantic value (in terms of aspect or modality)

4. Obviously, the “Subjunctive” category of the two Torres languages owes its name to very similar mood categories found in other languages (Noonan 1985: 91), notably Indo-European ones. This being said, as a principle, the observations made in this article must be understood as applying primarily to the TAM category specific to the Torres languages – hence the uppercase in its label, following the usage in Comrie (1976:10). My intention is not to make any general claim about the properties of a universal category *subjunctive* – supposing such a cross-linguistic category indeed exists (see Haspelmath 2007).

with some pragmatic property. In both cases, this property corresponds to a form of *pragmatic demotion* – lack of a specific illocutionary force for the Subjunctive, lack of focal status in the case of the Background Perfect – and in both cases, this demotion results in a form of clause dependency. While they are ultimately grounded in the pragmatic dimension of discourse, these two TAM-based strategies ultimately also affect the formal syntax of the sentence, as they constitute a routinized device for encoding clause subordination.

The following sections are organized as follows. Section 2 will provide a brief syntactic overview of the two Torres languages, and will pay special attention to overtly marked clause-linking strategies – whether subordination, coordination or verb serialization. Section 3 will then examine in detail the functional and formal behaviour of the Subjunctive, and Section 4 will be dedicated to the subordinating power of the Background Perfect.

## 2. Clause linking in the Torres languages: An overview

I will begin this study with an overview of the syntactic structures of the two Torres languages, with special focus on clause linking strategies.

### 2.1 Syntax of the simple clause

Like their Oceanic neighbours of Vanuatu, Hiw and Lo-Toga possess an accusative alignment system, and follow a strict SVO constituent order.

#### 2.1.1 Argument coding

Subjects take the form of noun phrases or free pronouns preceding the verb, and are not cross-referenced on the predicate itself. Likewise, direct objects usually leave the verb form unchanged (3a), except when they have human reference. In the latter case, the verb form becomes marked for transitivity (3b), and sometimes bears a suffix cross-referencing the object (3c):

- (3) a. LTG *Nēke na itē n' eñwe mē-he si.*  
1SG PRF<sub>1</sub> see ART house POSS-3PL PRF<sub>2</sub>  
'I've seen their house(s).'

- b. LTG *Nēke na ise kemi si.*  
1SG PRF<sub>1</sub> see:TR 2PL PRF<sub>2</sub>  
'I've seen you<sub>[+human]</sub>'

- c. LTG *Nēke na isi-he si.*  
1SG PRF<sub>1</sub> see:TR-3PL PRF<sub>2</sub>  
'I've seen them<sub>[+human]</sub>'

### 2.1.2 Tense-Aspect-Mood categories

Alongside its arguments, a well-formed verb phrase entails the presence of a marker coding for aspect, mood and polarity. These three parameters are subsumed under a single paradigm of portmanteau morphemes. For example, the marker labelled Complete (a postclitic *piti* in Hiw, a proclitic *mat* in Lo-Toga) simultaneously encodes aspectual meaning (completed event), modal meaning (indicative), and polarity value<sup>5</sup> (affirmative):

- (4) a. Hiw Sise mot̪ig piti.  
3PL sleep:PL CPLT  
b. LTG Nihe mat metur.  
3PL CPLT sleep  
'They've already slept.'

The category of tense properly speaking is not marked in these languages. Although the paradigm of verb modifiers should thus be designated, strictly speaking, as A-M-P markers (for "Aspect-Mood-Polarity"), throughout this chapter, I shall nevertheless continue to use the widespread abbreviation TAM (for "Tense Aspect Mood"), for the reader's convenience.

The two Torres languages possess sixteen formally distinct<sup>6</sup> TAM categories. The realis markers (see §4.1) include the Stative, the Imperfective, the standard Perfect, the Background Perfect, as well as the Complete, the Recent Perfect, and the Realis Negative. The unrealis categories (see §3.3) include the Future, the Prospective, the Potential, the Apprehensive, the Subjunctive, the Counterfactual, and the Unrealis Negative. Finally, two categories – labelled Aorist (see §2.2.1) and Time Focus – span the realis and the unrealis domains.<sup>7</sup>

The Aorist is a particularly polysemous category, found in the Torres<sup>8</sup> as well as several of the Banks islands to the south (François, in press). It covers several meanings, both realis and unrealis, including narrative, sequential, generic, prospective, imperative and conditional. A possible description of the Aorist would be to consider

5. The morphosyntax of the negation will be mentioned in §0.

6. Many of these TAM morphemes are morphologically complex, and sometimes discontinuous – as in the case of the Perfect *na...si* in (3). See also the discussion in §0.

7. See François (2003) for a detailed semantic analysis of a highly similar (and partly cognate) TAM system, that of the neighbouring language Mwotlap.

8. The morphology of the Aorist in the Torres languages is complex (François, in press). First, it is coded by a set of preverbal markers that vary for person and number (LTG 1sg *ke*, 2sg *we*, 3sg *ni...*); second, these preverbal markers are generally deleted in the presence of a free personal pronoun, in which case the surface form of the Aorist is simply Ø [see ex. (28b)]. In the present article, I will only mention the Aorist in the gloss when it is relevant to the discussion, otherwise the verb will simply be given as unmarked for TAM.

it as a "zero" verbal category that is underspecified with regard to tense, aspect and mood; this would account for both its great flexibility, and its compatibility with modality-bound dependent clauses (12). Interestingly, the Subjunctive [Hiw *on*, LTG *vé(n)*] can be analysed along similar lines – in terms of semantic underspecification – except that it is restricted to unrealis clauses (see §3). As we will see later, the two markers can be synonymous in certain contexts – compare (12) and (38) for modality-bound complement clauses; or (32f) and (35b) for the hortative. Yet even though the Aorist and the Subjunctive show a certain degree of functional overlap, the Subjunctive will be preferred when the semantic status of the subordinate clause is explicitly unrealis or generic.

### 2.1.3 Syntactic categories and their predicativeness

Another important characteristic of the Torres languages – and more generally of many of the area's languages (François 2005a) – is the diversity of parts of speech compatible with the predicate function. A predicate head<sup>9</sup> need not be a verb: it can be an adjective, a noun, a numeral, etc. For example, a nominal predicate takes the form of a simple noun phrase in a direct (zero) construction, with no copula – whether it be equational (type 'X is the N') or ascriptive ('X is an N').

- (5) Hiw Nine { ūekño-k }.  
3SG mother-1SG  
'She (is) my mother.'

When the subject is omitted, the result is a clause that consists of just a single noun phrase:

- (6) Hiw (Ø) { ne wake }.  
ART canoe  
'(It's) a canoe.' [DIRECT NOUN PREDICATE]

Several other word classes may also be directly predicative. This includes locative phrases – whether in the form of adverbs [e.g. the interrogative 'where' in (7)] or prepositional phrases [see *yō kōñ* in (54)] – as well as certain invariant words [e.g. the existential predicate 'not exist, lack' in (7)].

- (7) LTG Ne heñwēvot mino { evē }? – Nie { tategē }.  
ART knife my where 3SG NEG:EXIST  
'Where (is) my knife?' – 'It is not here.'

Direct predicativeness (Lemaréchal 1989; Launey 1994) constitutes an important property of parts of speech in the Torres languages, which will later prove crucial in the syntactic analysis of the Background Perfect (§4.2.2.2).

9. In Examples (5)–(7), the limits of the predicate phrase are indicated by curly brackets.

## 2.2 Subordination

Hiw and Lo-Toga possess a wide array of morphological devices for encoding the syntactic relations between a subordinate and a main clause. I will successively examine the coding of complement clauses (§2.2.1); conditional clauses (§2.2.2); relative clauses (§2.2.3); and adverbial time clauses (§2.2.4).

### 2.2.1 Complement clauses

The Torres languages have a quotative particle (Hiw *tom*, LTG *të*) for introducing direct reported speech. It can be used as the unique predicate of the clause, or in combination with a speech verb:

- (8) Hiw Tema-ne yuř-mi-e tom "Ye nēne?" Tom "Noke!"  
     father-3SG ask-TR-3SG QUOT who that QUOT 1SG  
     'Her father asked her [saying]: "Who was that?" [She said] "That was me!"'

The same quotative particle is used to introduce indirect speech. Therefore, despite its obvious origin as a quotative, it is better analyzed, synchronically, as a complementizer. Indeed it can combine not only with verbs of speech, but also with all sorts of verbs governing a clause complement:<sup>10</sup>

- (9) Hiw Noke tati mēnēg, noke ttōm tom ne gē kye.  
     1SG NEG steal 1SG think COMP ART thing my  
     'I didn't steal it, I thought (*that*) it was mine.'

- (10) LTG Ne ñwië ni holōq me, ni  
     ART devil AOR:3SG return hither AOR:3SG  
     itë tē nihe ve toge.  
     see COMP 3PL IPFV stay  
     'The devil came back, and saw (*that*) they were there'

If the complement clause is realis, its predicate is normally compatible with any realis TAM marker (Perfect, Stative, Imperfective...), with no particular restrictions. The same applies if the clause is semantically unrealis but is modally independent from the main clause. For example, a main verb meaning 'believe' would allow the complement clause to take essentially the same TAM markers as in an independent clause. As we shall see in §3.3.1, there are quite a few unrealis markers which correspond to this definition, for example the Potential (Hiw *ta*, LTG *si*):

- (11) LTG N' ige wē ne, nēke dōem tē nēke si gēn.  
     ART fish like this 1SG think COMP 1SG POT:AFF eat  
     'This sort of fish, I think I *can* eat.'

Conversely, certain types of predicates – typically, verbs of volition and manipulation – entail that the unrealis complement clause be bound to the main clause with respect to modality. In that case, the choice of TAM marking in the complement clause becomes essentially restricted to two possible categories: the Subjunctive [see (37)–(38) below] or the Aorist (12).

- (12) LTG Ne lie-k na ñih tē KE tun dē sa ñwil.  
     ART mind-1SG STAT want COMP AOR:1SG buy from M chief  
     'I'd like to buy it from the chief.'

Purpose clauses are also constructed along the same patterns (COMP + Aorist or COMP + Subjunctive): see (39)–(40) and (62)–(63) below. Once again, in this unrealis context, the Subjunctive and the Aorist are essentially equivalent (cf. §2.1.2).

The combination of the complementizer and Aorist markers has also grammaticalized, in Lo-Toga (but not in Hiw), into a TAM category in its own right, called the Prospective. Its meanings encompass the desiderative ('want to do'), the deontic ('should do', 'must do'), the prospective proper ('be about to do')...<sup>11</sup> Although it originally incorporates the complementizer *tē*, this Prospective marker can appear on the main predicate of an independent clause – as in (32c) below – which shows that it has lost any connection with clause dependency. This is also proven by the possibility of combining the Prospective (here *tē* *we* 'ProsP:2sg') with the complementizer *tē* in the same sentence:

- (13) LTG Tate pero tē nīke tē we hadit.  
     NEG:REAL long COMP 2SG PROSP 2SG be.initiated  
     [lit. It's not long before you're going to be initiated]  
     'You are soon going to follow the initiation rituals.'

The category of the Future is in turn a composite morpheme, which combines the Prospective (*tē* + Aorist) with the particle *ake* – see (15), (26), (32a).

### 2.2.2 Conditional clauses

Conditional clauses may again involve the same complementizer (Hiw *tom*, LTG *tē*), which is here translated 'if':

- (14) Hiw Tom ike gengon ñwō, ne ga tat qisi tiřtiř ike.  
     COMP 2SG AOR:eat first ART kava NEG:IRR hit:TR strong 2SG  
     'If you eat first, the kava won't have any strong effect upon you.'

10. This process, whereby the quotative particle has generalised its use to cover the whole functional array of a complementizer, is widespread in the area. The process may be compared to the typologically common process whereby complementisers originate in a verb of speech (see Heine & Kuteva 2002; Chappell 2008).

11. Both the morphology and the semantics of the Lo-Toga Prospective are narrowly similar to those of the Prospective in Mwotlap (François 2003: 218–257).

The conditional subordinator also displays longer forms which are derived from the complementizer. One thus finds the (semantically non-compositional) combination Hrw *tom* + *nwē* 'like' → *tom-nwē* or *tom-nwē-tom* meaning 'if' – see (49). Lo-Toga has exactly parallel forms, either morphologically transparent (*tē* + *wē* 'like' → *tēwē* [tēwē] 'if') or with a slight vowel change *tēwē* → *tewē* [təwē] ~ *tewē-tē* [təwētē] – see (15), (48).

Several TAM categories can be found in the protasis of a conditional sentence: Aorist; Subjunctive; Counterfactual (15):

- (15) LTG Tewētē TE not ne metē-ne si, nie tē n'  
       if CTFC<sub>1</sub> hit ART eye-3SG CTFC<sub>2</sub> 3SG FUT<sub>1</sub> 3SG  
       ake mēteqa ē!  
       FUT<sub>2</sub> blind OBL  
       'If they had hit his eyes, he would have become blind!'

We will see below (§3.5.2) that, while conditional constructions can make use of a conjunction, they are also regularly coded by the Subjunctive alone. This TAM marker is the only one capable of replacing a conditional conjunction.

### 2.2.3 Relative clauses

Relative clauses are marked by a variety of morphological devices. The role of relativizer can be played, in both languages, by the (polyfunctional) form *pe*:

- (16) LTG Noke tē ke vē k' itē ne gehuh  
       1SG PROSP 1SG go 1SG see ART coconut.crab  
       pe ve kerkur tēle nōk.  
       REL IPFV IPFV~crunch person there  
       'I'll go and have a look at that coconut crab WHICH devours people.'

The relativizer function can also be played by phonologically heavier forms; these combine several morphemes in ways that are not always semantically compositional. Thus one finds a relativizer Hrw *petom* ~ LTG *petē*, etymologically the combination {relativizer + complementizer} [also see (41) below]:

- (17) Hrw Sise mi nō-sa tiř nwute petom sise  
       3PL with POSS-3PL true place REL 3PL  
       toge ie yō nwēwōn.  
       stay:PL OBL:ADV LOC forest  
       'They have special places of theirs, WHERE they dwell in the forest.'

Lo-Toga also combines the relativizer *pe* with the comparative *wē* 'like' (→ LTG *pewē*), generally with virtual or generic referents (*whoever...*):

- (18) LTG Ni ole ne wuhe hi  
       AOR:3SG give ART potion DAT  
       heñwere pewē na mōo.  
       people REL STAT sick  
       'He provides medicine to WHO(EVER) is sick.'

In fact the form *wē* alone (without *pe*) can also serve as a relativizer in Lo-Toga – see (42). To sum up, the forms of the relativizer in Hrw are *pe* or *petom*; those in Lo-Toga are *pe*, *petē*, *pewē* or *wē*.

Finally, despite the wealth of these relativizers, it is also common for relative clauses to lack any formal subordinator, provided the status of the whole phrase as a dependent clause is visible on the verb's TAM marking. This ability to constitute a relative clause with no relativizer is attested only with two TAM categories, precisely those which form the topic of the following sections: the Subjunctive (§3.5.2), and the Background Perfect (§4.2.2.1).

#### 2.2.4 Adverbial time clauses

Adverbial time clauses are often formed with a noun meaning "time, moment": Hrw *tameřēn* ~ (take)timeřēn, LTG *mowe*. The time clause can then be construed as a relative clause (see Thompson & Longacre 1985: 179) – i.e. *when* = literally *the time in which...*

- (19) Hrw Ike yo-ie ti timeřēn pe kimiře nē  
       2SG see-3SG PRF time REL 2DU STAT  
       yumegov qē, tameřēn pe tekñwa te  
       young still time REL people from  
       yō vönyō ve tetaywō.  
       LOC village IPFV celebrate  
       'You met her (at a time) WHEN you both were still young,  
       as the villagers were celebrating.'

But it also commonly happens that the same word appears on its own, with no overt relativizer:

- (20) LTG Mowe ne tarepi ēke mat tēh pah,  
       time/when ART body canoe CPLT carve finish  
       pahvēn ge rak ne hēm' in.  
       then AOR:PL make ART outrigger its  
       'Once the body of the canoe is carved, [then] one makes the outrigger.'

It could be proposed to see *mowe* here still as a noun 'time' followed by a relative clause with no relativizer; however, such relative clauses, as mentioned in §2.2.3, are normally

restricted to two TAM markers. The presence in (20) of another TAM category (*mat* ‘Complete aspect’) calls for another syntactic analysis: namely, that the noun *mowē* has been grammaticalized into a subordinator ‘when’.<sup>12</sup>

In addition, Lo-Toga also has a genuine time subordinator *nonegē* ‘when, as’:

- (21) LTG *Nonegē nie ve vin-gē ne megole, ni*  
           as     3SG IPFV climb-APPL ART child   AOR:3SG  
*hur ne vete sise.*  
       sing ART song one  
     ‘As she was climbing with her baby, she began to sing a song.’

We shall see other cases where time clauses lack an overt subordinator, the relation of dependency being reflected only by the TAM marking on the verb: the Subjunctive (§3.5.2).

### 2.3 Coordination

The Torres languages make relatively little use of coordination, and generally prefer resorting to subordinating or serialising strategies.

Following a typologically common trend (Stassen 2000), the Torres languages usually form the equivalent of coordination between two noun phrases by using the comitative preposition *mi* ‘with’:

- (22) Hiw *tema-ne mi ūekña-ne*  
       father-3SG with/and mother-3SG  
       ‘his father WITH/AND his mother’

Quite originally, Lo-Toga has extended the use of this comitative preposition to coordination between any two phrases, including two prepositional phrases (23) or two clauses (24):

- (23) LTG *Noke na melekelake pi megole mēke, mi pi lēgie mēke.*  
       1SG STAT happy about child your and about wedding your  
       ‘I’m delighted about your baby, \*with/AND about your wedding.’
- (24) LTG *Ne ūwiē si dahia ē ne tēle, mi*  
       ART devil POT harm OBL ART person and  
*nihē si kur veriē ne tēle.*  
       3PL POT crunch also ART person  
       ‘Devils can harm people, \*with/AND they can even devour people.’

12. This pattern, whereby a noun meaning ‘time, moment’ grammaticalizes into a subordinator, is commonplace in the area. Mwotlap does the same with (*vēt*)*mahe* (François 2003: 26), as well as Bislama with *taem* < Eng. *time* (Crowley 2004: 188).

This functional extension of *mi* is unique to Lo-Toga, and is quite marginal in the language. It would be impossible in Hiw, where *mi* is still used strictly as a comitative preposition with a noun phrase. In order to coordinate two clauses, Hiw instead uses an adverb *pavēn* ‘then’:

- (25) Hiw *Timeřēn ēne, nine nē ūwotoy kē, pavēn n' uy ena nē teytoy.*  
       time   that 3SG STAT short little  
       then   ART hair her STAT plaited  
       ‘At that time, she was a little short, AND her hair was plaited.’

Other coordinate constructions include words for ‘but’ (Hiw/LTG *pa*), ‘or’ (Hiw *titom*, LTG *hitē*), or ‘because’ (Hiw [*uř*] *nēpe [tom]*, LTG *nawē*).

### 2.4 Verb serialization

Finally, this rapid overview of clause linkage in Hiw and Lo-Toga should mention, albeit briefly, verb serialization. Serial verbs in these two languages have two distinct forms.

The structure which is known in typology as *nuclear-layer* serialization (Foley & Olson 1985; Crowley 1987, 2002) consists in joining two verb radicals together with no intervening element, as if through a process of lexical compounding. The resulting “macro-verb” behaves in many regards as a single verbal unit, taking no more than one subject and one object:

- (26) LTG *Tē w' ake vese vahē noke ē ne iē igē.*  
       FUT<sub>1</sub> 2SG FUT<sub>2</sub> say show 1SG OBL ART name fish  
       [lit. You will say show me of fish names]  
       ‘You will teach me the names of fish.’

In this pattern of nuclear-layer serialization, the second verb modifies the first verb, both semantically and syntactically (Bril 2004; François 2004).

The Torres languages have also developed a pattern of *core-layer* serialization, whereby two verbs follow each other in a single clause, yet each one bears its own TAM marker (or at least the proclitic part, in the case of discontinuous markers). This TAM marker is normally the same for the two verbs:

- (27) LTG *Noke NA vēn NA vivdē SI l' ēñwe rōor.*  
       1SG PRF<sub>1</sub> go PRF<sub>1</sub> pray PRF<sub>2</sub> LOC house holy  
       ‘I went to pray in the church.’

This is an example of “concordant marking of tense-aspect-mood”, to use the terms in Aikhenvald (2006: 42).

The latter pattern is especially used when  $V_1$  is a verb of motion (*go, run...*) or of posture (*sit, stand...*). One of the derived uses of this serial structure, involving a posture verb in the  $V_1$ -slot, codes progressive aspect:

- (28) a. Ltg *Noke ve gel ve hiar ne heñwëvot mino.*  
          1SG IPFV stay IPFV seek ART knife my  
          [lit. I stay I search my knife]  
          'I am looking for my knife.'

This progressive construction involves either the Imperfective *ve* (cf. §4.1.1) as in (28b), or the semantically ‘neutral’ aspect called Aorist (§2.1.2). In this case, the very special morphology of the Aorist (fn.8 p.1) makes the serial pattern less easy to detect:

- (28) b. Ltg *Noke (Ø) gel ke hiar ne heñwëvot mino.*  
          1SG AOR stay AOR:1SG seek ART knife my  
          'I am looking for my knife.'

In all these cases, the sharing of arguments and of TAM marking – whether it occurs once or is repeated – clearly shows that we are dealing with serial verb constructions,<sup>13</sup> and hence with single clauses (Durie 1997; Bril 2004). As such, these structures do not illustrate patterns of clause linking strictly speaking, but rather linkage strategies between predicates.

### 3. The Subjunctive: In search of an illocutionary force

The preceding section showed the wide array of formal devices used by the two Torres languages to encode dependency relations between clauses and predicates, whether in the form of verb serialization, coordination, or subordination. Despite the wealth of these clause-linking devices, two TAM categories, the Subjunctive and the Background Perfect, present atypical behaviour: these two markers, and only these two, show a strong tendency not only to combine with subordinate clauses, but also to directly encode clause dependency, even in the absence of any subordinating device (see §1.2).

I shall detail these two cases successively: the Subjunctive in the present section, and the Background Perfect in Section 4.

13. The Resultative constructions of Lo-Toga share certain properties with these serial verb constructions, yet they must be analyzed as a different structure: see §0.

#### 3.1 Presentation

The Subjunctive was first exemplified in sentence (1), reproduced below:

- (1) Hrw Ne temët on tō yaqe me ñwë  
          ART devil SB JV go:SG appear hither like  
          ne, tekñwa voyi.  
          this people AOR:run.away  
          [lit. The devil would appear like this, people ran away]  
          '(Whenever) the devil appeared, people would run away.'

The behaviour of the Subjunctive is parallel in Hiw (form *on*) and in Lo-Toga (forms *vë ~ vën*).<sup>14</sup> One question arises: what exactly is the mechanism that makes this Subjunctive marker so intimately connected with subordination? Why is it that all other TAM categories – including the various irrealis markers – require the presence of overt subordinators, whereas the Subjunctive can easily do without them? Could one go as far as to consider this morpheme intrinsically endowed with the power of subordination?

The position I will adopt here is the following: the syntactic properties of the Torres Subjunctive, in terms of its ability to encode subordination, can be understood as an indirect consequence of fundamentally semantic properties: this marker codes an event as merely irrealis, *with no further specification of any illocutionary force*. This *modal and pragmatic indeterminacy* accounts for the inability of the Subjunctive alone to constitute well-formed utterances, and ultimately helps explain its strong tendency to trigger syntactic dependency between clauses.

#### 3.2 A note on irrealis sentences

An irrealis sentence involves the representation of a virtual situation which has no reality other than that of a mental construct in the speaker’s discourse. Unlike realis events, whose existence is a fact and which may therefore be recounted as such, an irrealis situation cannot simply stand on its own: in order to form a pragmatically well-formed utterance, it needs to be embedded in some form of secondary predication, be it a deontic predicate, an epistemic judgment, or a speech act of some sort.

14. Despite the formal difference between LtG *vë* [βe] ~ *vën* [βen] and Hrw *on* [ɔn], it is in fact likely that the two forms are cognate. According to regular vowel correspondences (François 2005b), they could reflect a proto-form \*βani, of uncertain origin. A link with Proto Oceanic \*pani ‘give’ is not implausible, although it raises semantic problems. The connection between give and subjunctives does not seem to be widely supported in other languages (see Bybee et al. 1994), and the etymology of English *if* (< OE *gif*), sometimes mentioned as connected to *giefan* ‘give’, is disputed.

For example, let us consider the state of affairs {BABY GET SICK}. When one refers to a realis event like (29), that state of affairs can easily be stated and provided with various semantic properties, such as time coordinates and truth value:

- (29) ENG Baby got sick again last week.

Conversely, the same state of affairs in an irrealis context (i.e. *the possibility that Baby gets sick at some point in the future*) will not be able to constitute, by itself, a complete utterance. Even the English sentence (30), which is syntactically complete and grammatical, appears to be an ill-formed utterance from the pragmatic point of view:

- (30) ENG Suppose Baby got sick.

A sentence like (30) is felt to be incomplete, as if waiting for the rest of the sentence in order to be interpretable.<sup>15</sup>

To use the terminology of Simon Dik's Functional Grammar, a sentence like (30) does little more than merely represent a possible State of Affairs – i.e. “the conception of something that can be the case in some world” (Dik 1989: 46). In order to constitute a well-formed utterance, such a virtual situation needs to be encapsulated within some type of higher-level linguistic operation – such as aspect and time operators that would provide it with the status of a “Possible fact”; or illocutionary force and modal values that would make it a pragmatically complete “Speech act”.

For example, the virtual state of affairs mentioned above could be incorporated within various forms of speaker-centered speech acts – e.g. apprehension, wish, prediction, etc.:

- (31) a. ENG *I fear* Baby might get sick.  
 b. ENG *I wish* Baby got sick!  
 c. ENG *[Given what I know, I hereby predict that]* Baby will get sick.

It may also take the form of a question, anchoring the modal center in the addressee (31d):

- (31) d. ENG *[According to you] will Baby get sick?*

It may also be encapsulated within a conditional structure, either as the protasis (31e–f) or as the apodosis (31g):

<sup>15</sup>. The pragmatic incompleteness of an English sentence like (30) is confirmed by historical evidence: in English-based Melanesian Pidgins such as Bislama or Tok Pisin, the imperative form *suppose* has grammaticalised into a subordinator *sipos/sapos* meaning ‘if’ (François 1997: 22; Mühlhäusler et al. 2003: 24; Crowley 2004: 189).

- (31) e. ENG *In case* Baby gets sick, *he will need to take this medicine*.  
 f. ENG *Every time* Baby gets sick, *he tends to recover within two or three days*.  
 g. ENG *If he goes out in that cold weather,*  
*[I hereby predict that] Baby will get sick.*

In all of these sentences, the virtual situation – which by itself has no pragmatic value – comes explicitly incorporated within a higher-level predication involving a specific speech act or modal attitude (prediction, wish, apprehension...). This is what makes them capable of forming a valid utterance, unlike (30) above.

### 3.3 Two types of irrealis markers in the Torres languages

These preliminary remarks about the nature of irrealis utterances should help understand the facts in Hiw and Lo-Toga. In each of these two languages, a semantically irrealis verb can be associated with two types of TAM markers: (a) modally specified markers, (b) a modally underspecified marker, the Subjunctive.

#### 3.3.1 Modally specified irrealis TAM markers

One set of irrealis TAM markers consists not only in representing a state of affairs as virtual; they also inherently convey a specific modal value and/or speech act (such as prediction, order, warning, etc.) within which this state of affairs is logically embedded. In a way, these modally specified morphemes could be described as semantically composite, as they combine the [+irrealis] feature with some other modal specification. It is therefore not surprising – following the reasoning in §3.2 – that they should be capable of forming pragmatically well-formed, complete utterances.

In Lo-Toga,<sup>16</sup> this first set of irrealis markers includes the affirmative Future *tē nāke* in (32a) and its negative counterpart *tat* in (32b); the Prospective *tē ni* in (32c); the affirmative Potential *si* in (32d) and its negative counterpart *tat ho* in (32e); the Aorist used for orders in (32f); the Apprehensional *mik* in (32g).

- (32) a. LTG *Nie tē n' ake metur l-eñwe mino.*  
 3SG:INDEP FUT<sub>1</sub> 3SG:S FUT<sub>2</sub> sleep LOC-house my  
*(I predict/promise...) 'He will sleep at my house.'*  
 b. LTG *Nie tat metur l-eñwe mino.*  
 3SG:INDEP NEG:IRR sleep LOC-house my  
*(I predict/forbid...) 'He won't sleep at my house.'*

<sup>16</sup>. The forms given in this paragraph are for Lo-Toga. Hiw has corresponding markers for all of them, except that it does not formally distinguish between the Future (32a) and the Prospective (32c).

- c. LTG *Nie tē ni metur l-eñwe mino.*  
       3SG:INDEP PROSP 3SG:S sleep LOC-house my  
       (I recount somebody else's desire...)  
       'He'd like to sleep/He's supposed to sleep... at my house.'
- d. LTG *Nie si metur l-eñwe mino.*  
       3SG:INDEP POT:AFF sleep LOC-house my  
       (I allow or state a factual possibility...)  
       'He can sleep at my house.'
- e. LTG *Nie tat ho metur l-eñwe mino.*  
       3SG:INDEP NEG:IRR POT:NEG sleep LOC-house my  
       (I state a factual impossibility...)  
       'He cannot sleep at my house.'
- f. LTG *Nie ni metur l-eñwe mino!*  
       3SG:INDEP AOR:3SG sleep LOC-house my  
       (I order/suggest...) 'Let him sleep at my house!'
- g. LTG *Nie mik metur l-eñwe mino!*  
       3SG:INDEP APPR sleep LOC-house my  
       (I present a situation as undesirable...)  
       '(I fear) he might sleep at my house!'

### 3.3.2 The Subjunctive, a modally underspecified TAM marker

In addition to these “modally specified” markers, the two Torres languages possess another irrealis marker with slightly different properties. This proclitic, which I label the Subjunctive, belongs to the same morphosyntactic paradigm as the TAM markers cited in (32a–g).

The reason for treating this morpheme separately is not morphological, but semantic. In itself, the Subjunctive provides the clause with no specific modality nor illocutionary force of any kind, and appears to convey the sole meaning [+irrealis]. To use Dik’s terms, it does nothing more than to represent a purely virtual *State of affairs*. It is therefore hardly surprising (following §3.2) that the Subjunctive alone is unable to form a pragmatically valid declarative sentence:

- (33) a. Hiw \*N' on mitiř yō-ñwe kye.  
           3SG SBJV sleep:SG LOC-house my  
       b. LTG \*Nie vēn metur l-eñwe mino.  
           3SG SBJV sleep LOC-house my  
       ('for him to sleep at my house'...)

A declarative sentence like (33a–b) would be felt to be truncated or unfinished, in a way very similar to (30) in English. This can be explained if one remembers that a virtual state of affairs can only form a complete sentence if it is embedded in a higher-level linguistic operation. While the various irrealis morphemes cited in (32a–g) incorporate

that linguistic operation intrinsically, this is not the case for the Subjunctive (33a–b), which remains *modally under-specified*.

This semantic property of the Torres Subjunctive entails an important corollary: its *high potential for syntactic dependency*. Due to its pragmatic incompleteness, a Subjunctive clause will need to hook on to some other clause or predication operator, in order to form a valid sentence. This essentially implies that the Subjunctive has a strong affinity with syntactic subordination – hence my choice for its name. In certain cases, this affinity means that the Subjunctive will combine with/be required by formal subordinators, in a way reminiscent of the subjunctives found in European languages. But quite often – and crucially for the topic of the present volume – the syntactic consequence will be that the Torres Subjunctive is capable of creating a relation of dependency between two clauses, *even in the absence of any specific subordinator*.

These issues will form the essentials of the discussion in §3.5. But before we turn to them, it is necessary to address the paradox of the hortative.

### 3.4 The special case of the hortative

The preceding paragraphs may have given the impression that the Torres languages make it impossible for an utterance to consist of a single clause marked as Subjunctive. Even though this may be indeed very close to the truth, there is in fact one exception to this generalization: the case of third-person hortatives.

When the speaker orders that an action be performed by the addressee, he will use an *imperative*. In the Torres languages this may be marked by the Aorist, or more often by the verb alone:

- (34) a. Hiw Tō me! ~ Wōt tō me!  
           go:SG hither AOR:2SG go:SG hither  
       b. LTG Vēn me! ~ We vēn me!  
           go hither AOR:2SG go hither  
       'Come here!'

When the person in control of the desired state of affairs is distinct from the addressee, the corresponding speech act, described typologically as a *hortative* (van der Auwera, Dobrushina & Goussov 2008), may also be coded by the Aorist, as in (32f) above. In addition, for *third-person hortatives*, the two Torres languages can also use their Subjunctive:

- (35) a. Hiw N' on mitiř yō-ñwe kye!  
           3SG SBJV sleep LOC-house my  
       b. LTG Nie vēn metur l-eñwe mino!  
           3SG SBJV sleep LOC-house my  
       '(I order/suggest...) 'Let him sleep at my house!'

This functional equivalence between the Aorist and the Subjunctive is also found with *third-person optatives*:

- (36) LTG *Ne teñwête vën toge mē-ke!*  
 ART peace SBJV stay with-you  
*(I wish) May peace be with you!*

This use of the subjunctive for hortatives or optatives is typologically common,<sup>17</sup> as witnessed by Latin *Veniat!* ‘Let him come!’ or *Pax sit semper vobiscum* ‘May peace be always with you’ (cf. Ernout & Thomas 1953: 239). However it seems to be at odds with the definition I gave of the Torres Subjunctive in §3.3.2, where it was stated that this marker does not convey any speech act value. If this is so, then where does the illocutionary force of these hortative or optative utterances find its source? And how is it possible that sentences such as (35a–b) and (36) are perfectly well-formed, while (33a–b) was ungrammatical?

The answer to this paradox does not lie with the Subjunctive itself: obviously, if hortative/optative modality were intrinsically built into this marker, then it should convey it in every sentence, and an utterance such as (33a–b) should be correct. This means we need to take seriously the only difference that distinguishes (33) from (35): the prosody – which is very roughly represented here by the punctuation. On the one hand, the prosodic contour of (33a–b), that of a declarative statement, results in the pragmatic incompleteness of the sentence. On the other hand, the prosody of (35a–b), which is characteristic of orders and exclamatory sentences – a high pitch plateau ending in an instant fall – makes the sentence grammatical.

In my interpretation, the particular suprasegmental profile of the sentence is the locus where the needed illocutionary force is lodged, and must be sought. The ungrammaticality of (33a–b) showed that the function of the Subjunctive, namely the mere representation of a virtual State of affairs, did not find enough support in the declarative modality to constitute a well-formed utterance. Conversely, what (35a–b) reveals is that an intonation typical of orders and exclamations, because it is markedly anchored in the speaker’s desires and emotions, is sufficient to provide that virtual State of affairs with the modal value and illocutionary force it needs to form a correct utterance.

Semantically, this formal asymmetry indeed makes sense. Such a mental construct as a virtual state of affairs can hardly be *stated* in any way; but it can still be represented in an emotional perspective – which is what exclamatory utterances tend to mimic. This contrast accounts, respectively, for the incompatibility between the Subjunctive and the declarative modality, and for its affinity with the intonation of orders and exclamations.<sup>18</sup>

17. See Noonan (1985: 54): “Main clause subjunctives tend to be used in modal, hortative, or imperative senses”.

18. A similar hypothesis was proposed in François (1997: 66) to explain why certain languages encode their imperative with some linguistic structures (noun phrases, infinitives, subjunctive clauses...) which would constitute an ill-formed declarative sentence. Despite their

In sum, (35) and (36) constitute no exception to the general principles outlined in §3.3.2, namely that an irrealis event can constitute a sentence if, and only if, it is involved in a modal predication of some kind. But while every other irrealis TAM marker in the Torres languages has an inbuilt illocutionary force that makes it well-designed for the formation of a valid utterance – cf. (32a–g) – this is not the case with the Subjunctive, which is under-specified in this regard. As a result, the only way for a Subjunctive verb to form a correct sentence, is to receive its illocutionary force “from outside”. Most of the time, this external source for the coding of modality will correspond to a different clause, that syntactically belongs outside the Subjunctive clause; this point will account for the strong ties of this marker with syntactic subordination (§3.5). As for (35a–b) and (36), they illustrate a more particular case, where the specific illocutionary force is lodged “outside” the verbal form strictly speaking, yet still has to be found within the formal limits of the clause itself: in its prosody.

All things considered, the functions of hortative and optative which are sometimes fulfilled by the Subjunctive do not contradict its earlier description as a *modally under-specified*, indeterminate irrealis marker.

### 3.5 From modal indeterminacy to syntactic subordination

In sum, the Subjunctive is the only irrealis TAM category of the Torres languages which does not inherently convey any modal value or illocutionary force. Unless it receives the latter from some modally charged intonation pattern, it is therefore unable to constitute a valid utterance on its own.<sup>19</sup>

The principal corollary of this description is the strong ties that exist between this irrealis TAM marker and the syntax of clause dependency. I will first review the various cases where the Subjunctive combines with a clause that is already marked formally as subordinate: complement clauses, relative clauses, conditional sentences, etc. In a subsequent section (§3.5.2), I will show that the presence of an overt subordinator is in fact not even necessary for the Subjunctive to be able to encode syntactic dependency between clauses.

#### 3.5.1 The subjunctive combined with overt subordinators

Quite often, the backgrounded clause is already marked as dependent by means of a subordinator of some sort. This is the case, for example, when a clausal complement is

morphological variety, these linguistic structures all share a similar semantic function: the representation of a virtual State of affairs. More recently, Nick Evans has addressed similar issues under the cover term “Insubordination” (Evans 2007).

19. This TAM marker corresponds to what Cristofaro (1998, 2003) calls a “deranked” verb form: that is, a form – of which the Italian Subjunctive would be another illustration – “that is structurally different from those used in independent declarative clauses” (Cristofaro 2008).

introduced by means of a complementizer (Hrw *tom*, LTG *tē*), after a verb of manipulation or expectation (see §2.2.1):

- (37) Hrw *Mařenage* *sa* *gatēt* *ti* *tekñwa* **TOM**  
chief their say DAT people COMP  
*ne* *veřoye* **on** *pa.*  
ART war SBJV finish  
[lit. The chief asked the people *that the war be stopped*.]  
'The chief asked his people to stop the war.'
- (38) LTG *Dege* *toge* *sēh* **TĒ** *ne* *gengēn*  
IINCL:PL stay wait COMP ART food  
*vē* *howse* *pah.*  
SBJV cooked finish  
'Let's wait till the food is completely cooked.'

The same formal structure {complementizer + Subjunctive} is used for purpose clauses, either with the same subject or with one different from the main clause,

- (39) Hrw *Sise* *myō* *ti* *ne* *töt* *ga* *ñot* **TOM** *sise*  
3PL pull PRF ART root kava one COMP 3PL  
*on* *ni* *yō* *gemoy.*  
SBJV drink LOC men's.house  
'They've pulled out a head of kava *so as to* drink it in the men's house.'
- (40) LTG *Hōr* *t'* *ō* *ñwule* *me* *vete* *mi* *hōr* **TĒ**  
3DU PROSP DU:S return hither place POSS 3DU COMP  
*nie* *vē* *menēwe.*  
3SG SBJV breathe  
'They are going back to their place for him to get some rest.'

Syntactic dependency may also be marked by a relativizer (§2.2.3). The Subjunctive is required when the relative clause is semantically unrealis and/or generic (cf. Eng. *whoever*):

- (41) Hrw *Tekñwa* **PETOM** *sise* **on** *tati*  
people REL 3PL SBJV NEG  
*voyi* *wřog.* *ne* *temēt* *quř-ise.*  
escape through ART Ghost crunch-3PL  
'All those who were unable to escape, the monster would devour them.'
- (42) LTG *N'* *ēve* *wē* *nihe* *vē* *vese* *hivi-ke,*  
ART thing REL 3PL SBJV say DAT-2SG  
*nike* *rōñtē* *urvē.*  
2SG listen properly  
'Whatever they may tell you, you must obey them.'

As we saw in §2.2.4, adverbial time clauses generally take the form of a relative clause hooked on the noun 'time, moment', with or without an overt relativizer. When the time reference of the subordinate clause is unrealis or generic, the Subjunctive is expected:

- (43) HIW *TAKETIMERËN* **PE** *ne* *tayō* **on** *mēt, tite*  
time REL ART person SBJV die IINCL:PL  
*tivig* *n'* *opē-ne.*  
bury ART body-3SG  
'When(ever) somebody dies, we bury their body.'
- (44) LTG **MOWE** *wē* *si* *tēle* *vē* *mōo, dege* *leklok* *mē.*  
time REL some person SBJV sick IINCL:PL help with.3SG  
'When(ever) somebody gets sick, we help them.'
- (45) LTG **MOWE** *kemē* *vē* *da-togin, nīke* *vēn* *me*  
time/when 1EXCL:PL SBJV be-ready 2SG go hither  
*dege* *ñwule.*  
1INCL:PL return  
'When we're ready, you can come here so we can go back together.'
- (46) LTG *Noke* *tē* *ke* *vēn* *ke* *tugtugerē* *remē* *mino*  
1SG PROSP 1SG go 1SG watch mother my  
**MOWE** *nie* *vē* *metur.*  
time/when 3SG SBJV sleep  
'I will watch my mother *when she's asleep*.'

An unrealis clause can be embedded within another unrealis clause, in which case the Subjunctive percolates throughout. (47) shows three instances of *vē(n)*: the first one (*vēn itē*) is due to the semantic status of the time clause as generic ('whenever'); the next two (*vē sōw vē lewō*) constitute a second level of subordination, being a complement clause within that time clause [see also (51) below]. Incidentally, the string /*vē sōw vē lewō/* is a serial verb construction, of the type that requires the repetition of the TAM marker (see §2.4):

- (47) LTG {**MOWE** *kemēm* *vēn* *itē* [**TĒ** *ne* *ho*  
time/when 1EXCL:PL SBJV see COMP ART leaf  
*in* *vē* *sōw* *vē* *lewō* *pe* *si* ],}  
its SBJV grow SBJV big already PRF  
*alē* *kemē* *ge* *lio.*  
then 1EXCL:PL AOR:PL dig.up  
'When(ever) we see that [the taro's] leaves have grown (and become) big,  
we dig it up.'

Finally, the protasis of conditional sentences (§2.2.2) constitutes another structure where the Subjunctive often combines with the subordinator ‘if’:

- (48) LTG TEWĒTĒ *ne liō vē ñih, níke si hue*  
           if ART mind:2SG SBJV want 2SG POT paddle  
           *o rōw vete qe ē.*  
           out out place deep OBL  
           ‘If you want, you can also paddle (your canoe) out there into the deep sea.’
- (49) HIW {TOMÑWĒTOM *se on vēn yō veñoye, s'*  
           if 3PL SBJV go:PL LOC war 3PL  
           *on qētñog ne tayō ne tayō on qēt}, sise*  
           SBJV kill:PL ART person ART person SBJV die:PL 3PL  
           *vīye n' opē-se me se mok erē qōr.*  
           take:PL ART body-3PL hither 3PL put on grave  
           ‘{ If/WHEN the population went to war, and many people were killed and died }, their bodies were then collected and deposited in stone graves.’

Note that the Subjunctive never occurs in the *apodosis* of such conditional sentences, because this is a section of the sentence which needs to have its own illocutionary force – as in (31g) above.

These Examples (37) to (49) all illustrate the strong links of the Subjunctive with subordinate structures. In each case, the Subjunctive verb phrase does no more than represent a virtual state of affairs which is, in itself, deprived of any inherent modal value. What then makes the clause interpretable, is its insertion – here via overt subordination – within a higher level predication, which is in turn specified for modality and illocutionary force.

### 3.5.2 The subordinating effect of the Subjunctive alone

Crucially, while the Subjunctive marker often combines with a subordinator, it turns out that it is also capable of creating an effect of syntactic dependency on its own. A clause marked as Subjunctive will spontaneously tend to develop a relationship of syntactic dependency with a matrix clause, even in the absence of any formal subordinator.

The most frequent case of “spontaneous” subordination is when the Subjunctive alone marks the protasis of a conditional sentence. Indeed, the conditional conjunctions ‘if’ – illustrated in §2.2.2 and in (48)–(49) – become optional when the verb is marked with the Subjunctive. In the majority of cases, the TAM morpheme is sufficient to encode the subordinate status of the clause:

- (50) LTG Nëke *vē vese tē ne genegone vē vēn,*  
           1SG SBJV say COMP ART war SBJV go  
           *ne genegone tē ni vēn.*  
           ART war PROSP 3SG go  
           ‘(If) I say that the war (must) go on, then the war will go on.’

- (51) HIW Ik' *on rōñ tom sē gē on ūak ti,*  
           2SG SBJV hear COMP some thing SBJV make PRF  
           *ike ta tōw ne wēt eye.*  
           2SG POT compose ART song OBL  
           ‘(If) you hear that some event has happened,  
           you can compose your song about it.’

- (52) HIW Ik' *on sēr-ie on yoqse, n' ēptigō nēne!*  
           2SG SBJV spear-3SG SBJV miss ART shame that  
           ‘(If) you try to spear him and you miss, then shame on you!’

Rather than hypothesizing a form of conjunction ellipsis, it is probably more accurate to suggest that the semantic dependency is inherently encapsulated in the modal morpheme itself.<sup>20</sup> Quite often, this leads to the impression that the Subjunctive marker itself is in fact a sort of conjunction meaning ‘if’. Consider for example the idiomatic phrase ‘if not’ (HIW *on tego*; LTG *vē tategē*):

- (53) HIW Tite *gon ne pēta, on tego, gon ne qēte.*  
           1INCL:PL eat ART yam SBJV NEG:EXIST eat ART taro  
           ‘We'll eat some yam; if not (=or else), some taro.’

The similarity of the Subjunctive morpheme with a conditional conjunction is not merely an effect of translation, but also appears to be a reality for the speakers themselves. This is clear, for example, in this sentence of HIW:

- (54) HIW On *yō kōñ, sise yō ñwutuye ne vti*  
           SBJV/if LOC night 3PL see only ART star  
           *ve yay ūē mesaye.*  
           IPFV shine on sky  
           ‘If at night, they would just watch the stars that shine in the sky.’

It is true that locative phrases – including prepositional phrases like *yō kōñ* ‘at night’ – may be used with the syntactic function of predicate (§2.1.3). However, this is always done in the form of a direct predicate, incompatible with any TAM marker.<sup>21</sup> Therefore, the combination of the subjunctive *on* with the phrase *yō kōñ*, rather than being seen as plain TAM marking – which would be grammatically abnormal here – would probably be better explained by a form of specialization of *on* as a form of (quasi)

20. A similar pattern of grammaticalisation can be found in some West Germanic languages. Thus in English, the modal auxiliary *should* in sentence-initial position takes up the function of a conditional conjunction: e.g. *SHOULD you be in Paris, call me* (see Van der Auwera & Plungian 1998: 98).

21. In other words, the part of speech LOCATIVE in these languages is “directly predicative”, but not “TAM-sensitive” (François 2005a: 192).

conjunction, similar to other conditional subordinators also present in this language (§2.2.2). Incidentally, this pattern is only attested in Hiw: Lo-Toga would have to use one of its genuine conjunctions here (*tewētē li qeñ* ‘IF [it were] at night’). This last point tends to confirm that (54) illustrates an extreme case in the evolution path of the Subjunctive, which Hiw has reached but not its close neighbour.

When a sentence-initial clause is marked by the Subjunctive alone, it can be ambiguous between a strict conditional reading – equivalent to the *if*-clauses of (48)–(49) – and a future or generic time interpretation – corresponding to the *when*-clauses of (43)–(47) above.

- (55) Hiw Ne ñwute on meyigeyige ttēñ, ike tēñ ar̄  
      ART place SBJV dark pitch 2SG grope seek  
      ñwutuye ne wōnaye.  
      just ART road  
      ‘(*IF/WHEN*) it's pitch dark, you just have to grope your way.’
- (56) LTG Ne tō vē ele gega wē nōk, tate  
      ART fowl SBJV crow always like this NEG  
      pero tē ne metave ni tōt.  
      long COMP ART morning AOR chop  
      ‘(*EVERY TIME*) the cock crows like that, (this means) day is almost breaking.’
- (57) LTG Deñwé'k noke vē ñwule, noke tē k' ole si vot.  
      today 1SG SBJV return 1SG PROSP 1SG:S take some stone  
      ‘Today (*WHEN*) I go back, I'll take some money with me.’

The proper interpretation will be given by the context. If the situation is expected to take place anyway – e.g. short-time visitors are expected to go back to their place sooner or later – it will translate as a *when* clause. But if the hypothesis is uncertain, then the topic clause will correspond to a conditional sentence proper. Obviously, the speakers get by perfectly well with this semantic ambiguity, and do not necessarily feel the need to disambiguate these situations, even though they do have the formal means to do so (see §2.2.2, §2.2.4).

The irrealis value of the Subjunctive does not only cover such time references as future and generic present. It is also found in past contexts – whether real or fictitious past, as in narratives – provided the event is presented as iterative:

- (58) Hiw Tekñwa on ñwuye me ton ne ūkove  
      people SBJV return hither from ART work  
      sa, s' on vēn wate me, se vēn se motřig.  
      their 3PL SBJV go:PL until hither AOR:3PL go:PL AOR:3PL sleep:PL  
      ‘(*EVERY TIME*) the group came back from their labour and reached home,  
      they would go to sleep.’

- (59) LTG Ne ñwiē vē ere nie vete'k, ni  
      ART Ogre SBJV hit:SG 3SG here AOR:3SG  
      wēl vēn wē nōk.  
      leap thither like this  
      ‘(*WHENEVER*) the Ogre tried to hit him, he would jump away like this.’

This is where sentence (1) – cited in §1.2 – would fit:

- (1) Hiw Ne temēt on tō yaqe me ñwē  
      ART devil SBJV go:SG appear hither like  
      ne, tekñwa voyi.  
      this people AOR:run.away  
      ‘(*WHENEVER*) the devil appeared, people would run away.’

This use of the Subjunctive in the expression of past iterative events in time clauses, paradoxical though it may be, finds its parallel in the Classical Latin “subjunctive of repetition” (Ernout & Thomas 1953: 400):

- LAT Id ubi dix-isse-t, hasta-m in fines  
      that when say-SBJV:PLUPRF-3SG spear-ACC to territory  
      eorum emitte-ba-t.  
      their throw-IND:IMPRF-3SG  
      ‘*WHEN(EVER)* he thus spoke, he would throw a spear into their territory.’  
      [Livy I, 32, 13]

Irrealis relative clauses are formed along similar lines. Compare (42) above with (60), where the subordinate status is exclusively coded by the mood marker:

- (60) LTG N' ēve nēke vēn alegōr tē tat rak,  
      ART thing 1SG SBJV forbid COMP NEG:IRR do  
      heñwere pah tē ge rōñtē.  
      people all PROSP PLS listen  
      ‘*Whatever I may ban people from doing, they will have to comply.*’

The presence of the article (*ne*) in (61) makes it clear that *mowe* is a noun meaning ‘time’ (rather than a subordinator, cf. §2.2.4), and that we are dealing here with an irrealis relative clause with no relativizer:

- (61) LTG ... vēn wahē ne mowe nēke vēn tēmetō.  
      go until ART time 1SG SBJV old.man  
      ‘... until the time (*WHEN*) I get old.’

Finally, while purpose clauses can include a complementizer as in (39)–(40) above, they may also do without any conjunction:

- (62) Hiw Noke tegegagyē ne megoye kye {n' on toge vařōñ}.
- 1SG IPFV~mislead ART child my 3SG SBJV stay quiet  
      ‘I'm amusing my baby (so) he keeps quiet.’

- (63) LTG *We tōw ne mon, we veñkē ne mesor*  
          AOR:2SG aim.at ART bird AOR:2SG let.go ART arrow  
          { *vē vēn vē qihē nie* }.  
          SBJV go SBJV bang 3SG  
          ‘You aim at the bird, then you shoot your arrow (so) it flies and knocks it.’

### 3.6 From clause dependency to lexical derivation

The pattern illustrated in (63), whereby a purpose clause can be coded by the Subjunctive *vē* alone, is the source of a process of reanalysis completed by Lo-Toga, but not Hiw. This process involves several steps leading to patterns of resultative (pseudo-) serialization, resultative compounding, and even causative derivation. I will conclude my analysis of the Torres Subjunctive by detailing the successive steps of this reanalysis. This will confirm the powerful affinity of the Subjunctive morpheme not only with clause dependency, but also with predicate binding, including an ultimate tendency towards the fusion of verb roots into one word.

Lo-Toga has developed a resultative construction that is clearly derived from the purposive subordination structure (63), yet with a tighter relationship between the two verbal heads, in a manner reminiscent of verb serialization. When a first dynamic event  $V_1$  (generally a verb of impact) results in a state  $V_2$ , then  $V_2$  is obligatorily marked as a Subjunctive. The structure {  $V_1 vē V_2$  } is particularly frequent in Lo-Toga:

- (64) LTG *Ole ne gi, ge tōt vē wureri,*  
          take ART kava AOR:PL chop SBJV small:PL  
          *ge gēt vē menō.*  
          AOR:PL chew SBJV soft  
          (*Procedural explanations about how to process kava, a woody plant*  
          *which is ground and brewed into a narcotic drink*)  
          ‘Take a branch of kava, mince<sub>[Aor]</sub> it small<sub>[Sbjv]</sub>, then chew<sub>[Aor]</sub> it soft<sub>[Sbjv]</sub>’

A sentence like (62) above unambiguously consisted of two distinct clauses: the main verb was immediately followed by its object (*the baby*), and the latter referent was repeated, in the form of a pronoun, as the formal subject within the subordinate purpose clause. Comparison between (62) and the two resultative constructions in (64) – respectively *tōt vē wureri* and *gēt vē menō* – shows similarities and differences. On the one hand, the underlying syntactic structures are identical: the subject of  $V_2$  coincides with the object of  $V_1$ . But on the other hand, (64) shows tighter structure than (62). Its two verbs are not separated by any noun phrase, be it the object of  $V_1$  or the subject of  $V_2$ ; the only morpheme

that divides  $V_1$  from  $V_2$  in each construction is the Subjunctive *vē*. Phonologically speaking, the strings {  $V_1 vē V_2$  } are uttered under a single contour with no internal pause, as if forming a single syntactic phrase.

The compactness of the constructions in (64) is confirmed by (65): if a noun phrase occurs, it is preferably postposed to the whole phrase {  $V_1 vē V_2$  } rather than inserted in between.

- (65) LTG *Dōr si gēt vē menō ne gi ne.*  
          1INCL:DU POT chew SBJV soft ART kava this  
          ‘We can chew this kava soft.’
- (66) LTG *Dege tē ge lōv vē ñwedōl ne*  
          1INCL:PL PROSP PLS call SBJV short ART  
          *iē tē Alex.*  
          your.name QUOT (name)  
          ‘We shall (pronounce shortly =) shorten your name to Alex.’

Functionally as well as formally, these strings {  $V_1 vē V_2$  } have a lot in common with serial verb constructions (§2.4), the only difference being that the TAM marking differs between  $V_1$  and  $V_2$ . Syntactically, this sequence of verbs behaves globally like a single, transitive macro-verb. In a way, it would even make sense to consider the whole string a single lexical unit (*gēt-vē-menō* ‘soften by chewing’; *lōv-vē-ñwedōl* ‘shorten’), as through a process of *lexical compounding*.

Arguably, the form *vē* in these compound forms has come to have a status of its own:<sup>22</sup> instead of coding the Subjunctive, it could be described here as a kind of “buffer” affix linking two verb roots together, with resultative meaning. This new analysis could result in an alternative transcription and gloss for (65):

- (65') LTG *Dōr si gēt-vē-menō ne gi ne.*  
          1INCL:DU POT chew-RESULT-soft ART kava this  
          ‘We can “soft-chew” this kava.’

Interestingly, Lo-Toga is the only language in north Vanuatu that has developed this pattern of resultative structure, using a buffer morpheme like *vē*. All its neighbours – including Hiw – would simply construct their resultative macro-verbs by resorting to a simple pattern of nuclear-layer serialization (François 2004, 2006). Thus, the equivalent of (65) in Mwotlap would be *kuy madamdaw na-ga/chew soft ART-kava/*, with nothing between the two verb radicals.

22. Note that the variant *vēn* is never attested in these new structures, which in other words tends to confirm that the Subjunctive marker has adopted a new grammatical status here.

While sentences like (64)–(66) are still somewhat ambiguous and compatible with more than one interpretation, some other examples provide an even clearer case for a compounding analysis. This is especially true when the first verb before *vë* is the dummy auxiliary *da* ‘do’ (also ‘be’), which does not exist as an independent verb. The string *da-vë-* thus serves as a productive prefix in Lo-Toga for the formation of causative (transitive) verbs out of stative (intransitive) verbs or adjectives (Table 1).

**Table 1.** Resultative compounds of Lo-Toga, incorporating the Subjunctive/Resultative morpheme *vë*

SIMPLE VERB/ADJECTIVE		RESULTATIVE COMPOUND		
<i>menō</i>	‘soft’	→	<i>gët-vë-menō</i>	‘soften by chewing’
		→	<i>qihih-vë-menō</i>	‘soften by grinding’
<i>ñwedöl</i>	‘short’	→	<i>lōv-vë-ñwedöl</i>	‘shorten (a name)’
<i>mōo</i>	‘sick’	→	<i>da-vë-mōo</i>	‘make s.o. sick, sicken’
<i>mēmerie</i>	‘painful’	→	<i>da-vë-mēmerie</i>	‘hurt (body part)’
<i>luwō</i>	‘big’	→	<i>da-vë-luwō</i>	‘make bigger, enlarge’
<i>hemrē</i>	‘laugh’	→	<i>da-vë-hemrē</i>	‘talk playfully, joke’
<i>duwér</i>	‘false’	→	<i>da-vë-duwér</i>	‘pretend’
<i>rōor</i>	‘holy’	→	<i>da-vë-rōor</i>	‘consecrate, baptize’

- (67) LTG *Temētrōñ tat ho da-vë-mōo ne tēle.*  
healer NEG:IRR POT:NEG do-RESULT-sick ART person  
'Healers cannot make people sick.'
- (68) LTG *Ne ri ñwēl na deda-vë-mēmerie*  
ART top.of reef STAT ITER-DO-RESULT-painful  
*ne teplē tēle.*  
ART foot person  
'The surface of the coral reef hurts the feet.'
- (69) LTG *Tate hehu da-vë-rōor nihe që.*  
NEG bathe do-RESULT-holy 3PL still  
[lit. (one) has not bathed consecrated them yet]  
'They haven't been baptized yet.'

Once again, these examples are open to two morphological analyses. It is still possible to consider them compoundings between two lexical roots (*da* ‘do’ + *mōo* ‘sick’), hence the gloss /do-RESULT-sick/. But due to the relative productiveness of the process, and the low semantic specificity of the first auxiliary, it would be equally accurate to speak synchronically of a process of lexical derivation that actually combines a single lexical

unit (*V<sub>2</sub>*) with a CAUSATIVE prefix *davë-*. In the latter case, one could transcribe (67) as *davë-mōo* and gloss it /caus-sick/.<sup>23</sup>

The historical and/or logical processes outlined here can be described as a series of morphosyntactic reanalyses. Starting from a clear pattern of subordination between two clauses, each step corresponds to a tighter relationship between the verbs of each clause, and ultimately results in a specialized pattern of causative derivation (Table 2).

**Table 2.** From biclausal purpose subordination to causative derivation: the binding power of the Subjunctive

Ex.	SYNTACTIC ANALYSIS	ROOTS	VERBS	CLAUSES
(40)	<i>V<sub>1</sub></i> = main clause <i>V<sub>2</sub></i> = dependent purpose clause, with subordinator	2	2	2
(62)	<i>V<sub>1</sub></i> = main clause <i>V<sub>2</sub></i> = dependent purpose clause, without subordinator	2	2	2
(64)	<i>V<sub>1</sub></i> = first action in resultative (quasi) serialization <i>V<sub>2</sub></i> = resulting state in resultative (quasi) serialization	2	2	1
(65)	<i>V<sub>1</sub></i> = first radical in resultative compound verb <i>V<sub>2</sub></i> = second radical in resultative compound verb	2	1	1
(67)	<i>V<sub>1</sub></i> = (dummy verb) > causative prefix <i>V<sub>2</sub></i> = stative verb, input of causative derivation	1	1	1

### 3.7 The Subjunctive: Summary

The various functions of the Subjunctive in the two Torres languages are summarized in Table 3.

**Table 3.** The narrow ties between the Subjunctive and clause dependency: A summary

SYNTAX	FUNCTIONAL VALUE	EXAMPLES
no subordination	hortative & optative (3sg)	(35)–(36)
	modally-bound complement clauses	(37)–(38)
combines with subordinators	purpose clauses	(39)–(40)
	irrealis & generic relative clauses	(41)–(42)
	irrealis adverbial time clauses	(43)–(47)
	irrealis conditional protases	(48)–(49)
directly encodes subordination	irrealis conditional protases	(50)–(52)
	⇒ [Hiw] reinterpreted as conjunction if	(53)–(54)
	irrealis & generic adverbial time clauses	(55)–(59)
	irrealis & generic relative clauses	(60)–(61)
	irrealis purpose clauses	(62)–(66)
	⇒ [LTG] resultative compounding	(65)–(69)
	> causative derivation	

23. This prefix has thus, in function, replaced the Proto Oceanic causative prefix \**paka-*, which has essentially left no trace in the two Torres languages.

#### 4. The Background Perfect: In search of a focus

The TAM category I propose to label “Background Perfect” offers a broadly similar, yet quite distinct illustration of the phenomenon just discussed with the Subjunctive. The general mechanism behind the two patterns is the same: the semantic and pragmatic identity of a TAM marker makes it particularly prone to the syntactic coding of clause dependency. Nevertheless, the case of the perfect is sufficiently different to warrant a section of its own.

The question addressed here is the following: how can the Background Perfect marker (*ve... si*) clearly form a subordinate – relative – clause in a sentence like (2), and yet do without any overt subordinator? What is there in its makeup that renders it syntactically different from other realis categories, and especially different from the regular Perfect?

- (2) LTG Ne *gehuh* ve kerkur tèle si mat mêt.  
          ART coconut.crab BKPF<sub>1</sub> ITER~crunch person BKPF<sub>2</sub> CPLT die  
          [lit. The coconut crab *has devoured people* has died.]  
          'The coconut crab (*who*) *had devoured people* was dead.'

Once again, I shall argue that the syntactic power of this marker must ultimately be understood as an outgrowth of its main functional property, namely, its ability to mark the informational status of its predicate as *presupposed*. Due to this form of *pragmatic demotion*, the predicate phrase thus marked needs to search for an external focus of information, which will typically result in a syntactic relation of dependency between clauses.

#### 4.1 The two perfects and sentential focus

Among the various TAM categories that can denote a realis event (§2.1.2), the two Torres languages have a Stative, an Imperfective, and two perfects. I will briefly present the first two of these TAM markers, before I turn to the difference between the last two which are derived from the former.

#### 4.1.1 Stative vs. Imperfective

The Stative [Hiw nē(gē), LTG *na*] is followed exclusively by stative predicates, that is, stative verbs and adjectives:

- (70) LTG Ne *vavetēme mi kemī na deřēňha*.  
          ART language POSS 2PL STAT difficult  
          'Your language is difficult'

The only way for a semantically dynamic verb to be compatible with this marker is to first be converted into a habitual (and therefore stative) predicate, by means of reduplication:

- (71) LTG \*(*Néke* na si.) → *Néke* na *sesi*.  
           \*1SG   STAT walk       1SG   STAT ITER~walk  
           \*(I walk.)           → 'I usually walk, i.e. I am a (good) walker.'

As for the Imperfective (Hiw/LTG *ve*),<sup>24</sup> it encompasses two aspectual values (cf. Comrie 1976): the progressive (72) and the habitual (73):

- (72) LTG *Remē mē ve kerē*.  
           mother his IPFV weep  
           ‘His mother is/was weeping.’

(73) LTG *Nihe ve lōv nie tē* “*Temētrōñ*”.  
           3PL IPFV call 3SG QUOT Healer  
           ‘People call him “Healer”’.

The same Imperfective *ve* also takes part in several progressive structures based on verb serialization { *ve Posture verb V*, + *ve Action verb V*, }: see §2.4, ex. (28a).

Verbs that are lexically stative (including adjectives) are sometimes found to combine with the Imperfective, in which case they take on a dynamic reading:



However, setting aside these rare cases, it is generally true that the Stative and the Imperfective tend to target two different sets of verbs, respectively stative and dynamic. Obviously this makes it difficult to carry out any extensive comparison of these two TAM markers. But as we shall now see, the situation is totally different for the two perfects that are derived from them.

#### 4.1.2 The two perfects

I now turn to the two perfects of the Torres languages, which will form the heart of the following discussion: the regular Perfect (*Hiw nē...ti/LTG na...si*) and the Background Perfect (*Hiw ve...ti/LTG ve...si*).<sup>25</sup>

24. Beside the widespread form *ve* [βə], Lo-Toga also possesses a rare variant *me* [mə]; likewise, *me...si* constitutes a (rare) variant of its Background Perfect *ve...si*. Incidentally, there is no reason to suspect any etymological connection between the element *ve* [βə] of the Imperfective and the Lo-Toga form of the Subjunctive *vē* [βə]: they are two unrelated morphemes.

25. Unlike Lo-Toga where the contrast is systematically coded, Hiw is problematic in that it treats the two proclitics – respectively *n̥e* and *ve* – as optional (see Table 4). Quite often, a

Morphologically speaking, one may say that these two perfect markers show a straightforward correspondence with the Stative and the Imperfective, as they simply consist in the combination of the latter with the postclitic \**ti*.<sup>26</sup> However, the clitic \**ti* only occurs in combination with TAM markers, with various semantic effects, and cannot be assigned any stable meaning unto itself. It is therefore methodologically safer – and probably more realistic from the speaker's point of view – to consider each compound TAM marker as a single meaningful morpheme, albeit a discontinuous one. As a result, while the form *na* alone was glossed STAT(IVE), I shall gloss the sequence *na...si* as PRF<sub>1</sub>...PRF<sub>2</sub>, with no attempt to arrive at a compositional analysis.<sup>27</sup> As for the semantic processes that may have led to the creation of these compound forms, that is a matter for history, and goes beyond the limits of the present study.

Considered from a purely semantic angle, the two TAM categories under consideration are synonymous, as they both correspond to the typological definition of the *perfect* aspect. They represent a realis event insofar as it is complete, and place the cursor in the resultant state that follows that event.

- (75) a. LTG *Kemëm na gil o si ne keka tekèle.*  
1EXCL:PL PRF<sub>1</sub> dig out PRF<sub>2</sub> ART yam some  
'We have dug out a few yams.'
- b. LTG *Ne keka tekèle kemëm ve gil o si.*  
ART yam some 1EXCL:PL BKPF<sub>1</sub> dig out BKPF<sub>2</sub>  
'(These are) a few yams we have dug out.'

Because they both point to the resultant state that follows the final boundary of a completed state of affairs, they are equally compatible with stative and with dynamic predicates. This contrasts with the Stative and the Imperfective, which tend to combine with distinct sets of verbs – stative vs. dynamic – as we saw earlier (§4.1.1). Thus, while the

perfect predicate will be tagged by the postclitic *ti* alone – as in (19) or (39) – blurring the contrast between the two perfects. This is why the present section will mainly cite examples from Lo-Toga, where the phenomenon is much more conspicuous. This being said, when the proclitics of Hiw are overtly marked – as in (76) or (79) – they do conform to the same principles as in Lo-Toga.

26. To be precise, Lo-Toga alternates between two allomorphs: an assibilated form *si* (< \**ti*), and an elided form *t'* [t] when preceded or followed by a vowel – see (80), (87), (88). Here I lump the two synchronic allomorphs together under the underlying (and ancestral) form \**ti*, for the sake of discussion.

27. I adopted similar methodological principles for the analysis of discontinuous TAM markers in Mwotlap (François 2003: 30 sqq, 343). Incidentally, most of the compound forms of Mwotlap involved postclitic *tō* [tu], with which the Torres form *ti/si* is cognate.

dynamic verb *gil o* 'dig out' is incompatible with the Stative *na* (→ \**na gil o*), it can perfectly take the standard Perfect which is derived from it (→ *na gil o si*).<sup>28</sup>

Yet, even though the two perfects may be said to be synonymous in terms of their aspectual semantics, they are not functionally equivalent, and in fact occur in distinct contexts. The difference between these two TAM categories is best defined in pragmatic terms, by contrasting the manner in which they organize the informational hierarchy within the sentence: to use the terms of Lambrecht (1994: 52), the standard Perfect puts its predicate under the scope of the *assertion*, whereas the Background Perfect explicitly encodes its status as a *pragmatic presupposition* (Table 4). This use of TAM markers for coding informational hierarchy is typologically original.

Table 4. Hiw and Lo-Toga have two Perfects; their difference lies in the pragmatic status of the predicate

	HIW	LO-TOGA	PRAGMATIC STATUS OF PREDICATE
(Standard) Perfect	( <i>në</i> )... <i>ti</i>	<i>na... si</i>	asserted/foregrounded
Background Perfect	( <i>ve</i> )... <i>ti</i>	<i>ve... si</i>	presupposed/backgrounded

#### 4.1.3 When TAM markers encode informational hierarchy

The regular Perfect (Hiw *në...ti*, LTG *na...si*) represents the predicate as a realis perfect event and it places it under the scope of the sentence's informational focus. This is the pragmatically unmarked situation, where the syntactic center of the sentence coincides with its pragmatic center in terms of assertion – as in (75a) or (76a):

- (76) a. Hiw *Ike ttöm tom ne tiř mon, pa*  
2SG think COMP ART true bird but  
*tego. Në ūkak ti.*  
NEG:EXIST PRF<sub>1</sub> make PRF<sub>2</sub>  
'You could think it's a real bird, but far from it. (*Somebody*) made it.'

As for the Background Perfect (Hiw *ve...ti*, LTG *ve...si*), it also construes a realis perfect predicate, but explicitly specifies its informational status as pragmatically presupposed, i.e. defocused. Crucially, a predicate phrase marked with the Background Perfect (henceforth "BkPf"), due to this backgrounded status, cannot constitute a well-formed utterance on its own:

- (76) b. Hiw \**Ve ūkak ti.*  
BkPF<sub>1</sub> make BKPF<sub>2</sub>  
\*{ (*somebody*) made it... }<sub>[background]</sub>

28. This freedom of actionality combinations provides further support for the view explained above, that the two perfects should not be analyzed compositionally, but as (discontinuous) TAM markers in their own right, with specific properties.

In contrast to (76a) *nē ūk ti*, a sentence like (76b) would be deemed incomplete. This is because an utterance, in order to be pragmatically valid, needs to contain at least some new, assertional information.<sup>29</sup> Insofar as the BkPf tags a predicate phrase as presupposed, it makes it unable to constitute a correct utterance by itself; in order to be interpretable, the sentence needs some other constituent with which the pragmatic assertion can be identified.

Occasionally, the background status applies to the whole clause (i.e. the predicate with its arguments and complements), which is then entirely marked as presupposed. This is what happens, for example, when the speaker refers back to an event that is already known to the addressee, as a reminder. Thus compare the regular Perfect of (77a), where the whole clause is fully new, and the Background Perfect of (77b), where it only serves as a reminder of an already known fact:

- (77) a. LTG *Sesē na hag si!*  
           your.sister PRF<sub>1</sub> sit PRF<sub>2</sub>  
           ‘Hey! { Your sister has given birth! }<sub>[focus]</sub>’
- b. LTG *Sesē ve hag si: ne teñwēn hitē ne leqēvine?*  
           your.sister BKPF<sub>1</sub> sit BKPF<sub>2</sub> ART male or ART female  
           ‘{ Your sister has given birth (*as we know*): }<sub>[background]</sub>  
           { is it a boy or a girl? }<sub>[focus]</sub>’

(77b) could be described as a case of clause topicalization.<sup>30</sup> The event marked as Background Perfect has no informational value in itself, that would allow it to form an utterance on its own; rather, it is used as a reminder to help the addressee interpret the focal part of the sentence (in this instance, the question).

#### 4.1.4 How many clauses?

As is typically the case for topic–focus structures, the syntactic relationship between the two clauses in (77b) is still loose. While it does illustrate a form of dependency, it does not form subordination in the strict sense of the term. Most of the time, however, the Background Perfect is involved in much more tightly bound structures than this.

29. See Givón (1984: 241), Tomlin (1985), Lambrecht (1994: 60).

30. Other strategies for clause topicalization have been observed, for example, with the “background topic clauses” found in Chuave, a language of Papua New Guinea (Thurman 1979, cited by Givón 1990: 870). Clause topicalization is a common phenomenon in North Vanuatu, but in the neighbouring Banks languages, it involves the use of deictics rather than of TAM strategies (François, in prep.).

As we shall see in §4.2, the presupposed predicate quite often involves genuine subordination, e.g. a relative clause:

- (78) LTG *Lōwie ē leqēvine meke { nie ve rak*  
           thanks OBL woman your 3SG BKPF<sub>1</sub> make  
           *si ne tōtōgalē }.*  
           BKPF<sub>2</sub> ART picture  
           ‘Thanks to your wife {(who) drew the pictures }<sub>[background]</sub>’

One ambiguous case, however, is when the sentence apparently consists of a single predicate: this happens especially in contrastive focus sentences like (79).

- (79) Hiw NOKE ve tot ti.  
           1SG BKPF<sub>1</sub> carve BKPF<sub>2</sub>  
           ‘I carved it!’ (not you...)

The predicate here (*ve tot ti*) is the presupposed segment of the sentence, whereas the focal part corresponds to its grammatical subject (*noke*). In fact the sentence’s structure is quite parallel to its English counterpart, including the contrastive focal stress that affects the subject phrase, with the same pragmatic implications. All these arguments tend to suggest that (79), just like its English translation, consists of just one syntactic clause, with no possibility to speak here of clause dependency. If this were true, then we would need to temper the claim that the pragmatic mechanism of the Background Perfect almost systematically goes along with subordination. In doing so, one would have to admit that the pragmatic properties of the BkPf sometimes trigger clause dependency as in (78), but sometimes operate on a purely pragmatic level, with little incidence on the syntactic structures, as in (79). This would also challenge the statement made earlier – on the subject of (76b) – that a main clause cannot stand alone if it is marked with the Background Perfect.

In fact, we shall see below (§4.2.2.2) that the structural similarity between Lo-Toga and English in (79) is an optical illusion. It will appear that (79), like all contrastive focus patterns in the Torres languages, is best analyzed as consisting of not just one, but two distinct clauses. In doing so, I will show that the Background Perfect does not only affect the pragmatic interpretation of the sentence in terms of informational hierarchy, but also has a syntactic impact, in creating a genuine subordination relation between predicates.

#### 4.2 From pragmatic presupposition to syntactic subordination

The following pages will illustrate in more detail this syntactic affinity between the Background Perfect and clause dependency. I will first show cases where the two perfects combine with overt subordinators (§4.2.1). I will then show that the BkPf alone may in fact suffice to generate clause dependency and subordination, without requiring

any further formal device (§4.2.2). The special syntax of contrastive focus structures will be examined in §4.2.2.2.

#### 4.2.1 *The two perfects and overtly marked subordination*

The semantic principles exposed in §4.1.3 for main clauses are equally true for those clauses which are formally marked as dependent by means of an overt subordinator. Thus, the regular Perfect will be used whenever the subordinate clause falls under the scope of the assertion. This is the case, in general, for complement clauses attached to speech or thought verbs:

- (80) LTG *Nēke dōem { t̄e ne ñwiē na kur*  
           1SG think COMP ART Ogre PRF<sub>1</sub> crunch  
           *nike pe t̄ }*.  
           2SG already PRF<sub>2</sub>

'I thought (that) the Ogre had already devoured you.'

Regular Perfects are also found in the protasis of certain conditional clauses:

- (81) LTG { TEWĒTĒ ne temēt na öla nīke si },  
       if ART ghost PRF<sub>1</sub> take:TR 2SG PRF<sub>2</sub>  
       *Temētrōn tē n' ake vēn hēr ē nīke Pene.*  
       Healer FUT<sub>1</sub> 3SG FUT<sub>2</sub> go find OBL 2SG Hell  
       'If the ghosts kidnapped you, the Healer would come and find you in Hell.'

Conversely, if a subordinate clause refers back to an already established event, then the Background Perfect will be required. This is especially true of restrictive relative clauses, whose function is precisely to point to a background element to help the addressee track referents:

- (82) LTG *ne revrev PE nēke ve hur si /*  
       ART song REL 1SG BKPF<sub>1</sub> sing BKPF<sub>2</sub>  
       (*?? pe nēke na hur si*)  
       REL 1SG PRF<sub>1</sub> sing PRF<sub>2</sub>  
       'the song { which I sang }<sub>[background]</sub>'

- (83) LTG *Ne līlie { PE nie ve durlue si }*  
       ART cave REL 3SG BKPF<sub>1</sub> drill BKPF<sub>2</sub>  
       *ve taqe wahē me deñwē'k.*  
       IPFV lie until hither today  
       'The cave { which he broke open }<sub>[background]</sub> still exists today.'

In each of these two sentences, the relative clause is unambiguously marked as subordinate by its relativiser *pe* (§2.2.3). As for the BkPf, it arguably operates on the pragmatic level, by providing its predicate with a background status.

The regular (assertive) Perfect is thus extremely rare in relative clauses. This configuration does occur however, in exceptional cases, when the informational focus is in fact located within the relative clause. Example (84) provides an illustration of this non-standard situation:

- (84) LTG *Heñwere pah tat lölmerēn ē.*  
       people all NEG:IRR know OBL:ADV  
       *Heñwere { wī na huqe wereño si }*  
       people REL PRF<sub>1</sub> initiated only PRF<sub>2</sub>  
       nihe ve lölmerēn ē.  
       3PL IPFV know OBL:ADV  
       'Not everybody would know (these things).  
       Only { those who've been initiated }<sub>[focus]</sub> know<sub>[background]</sub>'

Functionally, a mere repetition of the previous sentence, with no informational weight is what, syntactically, forms the main clause (*nihe ve lölmerēn*) of the whole sentence. Conversely, the sentence's assertion is located in the relative clause, which exceptionally takes the regular Perfect rather than the Background Perfect.

A sentence such as (84) tends to show that the use conditions of the two perfects in relative clauses do not obey a strict formal rule, whereby all relative clauses would mechanically take the Background Perfect. Rather, the choice of TAM marker remains a functionally productive device, based on the informational hierarchy chosen by the speaker in organizing his utterance.

#### 4.2.2 *The subordinating effect of the Background Perfect on its own*

In sum, it would be exaggerated to say that all relative clauses – let alone all subordinate clauses – require the Background Perfect: this is only the case for those clauses which are pragmatically presupposed. Now, if we narrow down our observations to the latter configuration, an important point remains to be made. Unlike all other realis markers, the Background Perfect allows a subordinate clause to dispense with any formal subordinator, as though it were sufficient *per se* to code for clause dependency. This, as we shall see now, is especially the case with relative clauses, and the focus cleft constructions which are derived from them.

##### 4.2.2.1 *Relative clauses*

While the BkPf is occasionally found to combine with an overt relativizer – see (82)–(83) – the most frequent pattern is for perfect relative clauses to dispense with a formal subordinator, and be simply marked by the BkPf alone (see also (78) above).

- (85) LTG *li megage { ve pah si }*  
       LOC month BKPF<sub>1</sub> finish BKPF<sub>2</sub>  
       'last month' [lit. in the month {(which) has finished}]<sub>[background]</sub> ]

- (86) Hiw Ike peon sawe-vog ne temēt { tekñ̄wa ain ve  
 2SG FUT dance-APPL ART headdress people other BKPF<sub>1</sub>  
 rak ti }.  
 make BKPF<sub>2</sub>  
 'You shall dance with a headdress { other people will have made }<sub>[bkg]</sub>'

A superficial look at (86) could suggest a comparison with the syntax of zero-marked relative clauses in English, which happens to be parallel here. Two differences must however be noted.

- Contrary to English, zero-marked relative clauses in Torres languages are allowed whatever the function of the antecedent within the subordinate clause. Thus while English allows a zero-marked clause in (86) where the relativized NP is an object, it does not in (85), where it is a subject. The Torres languages are less constrained than English in this regard.
- Contrary to English, zero-marked relativization in the Torres languages is only allowed in combination with certain specific TAM markers, the Background Perfect and the Subjunctive. The Torres languages are more constrained than English in this regard.

We can now account for Example (2), which was quoted in §1.2:

- (2) LTG Ne gehuh { ve kerkur tēle si }.  
 ART coconut.crab BKPF<sub>1</sub> ITER~crunch person BKPF<sub>2</sub>  
 mat mēt  
 CPLT die  
 [lit. The coconut crab { has devoured people }<sub>[background]</sub>  
 { has died }<sub>[focus]</sub>]  
 'The coconut crab (who) was devouring people had died.'

On the face of it, (2) is a sequence of two clauses taking the same subject, with no formal dependency marker between the two clauses. Only the nature of the Background Perfect, and its ability to defocus its own predicate, makes it clear here which clause is subordinate, and which is the main clause of the sentence. It must also be noted that – setting aside the case of the Subjunctive (§3.5.2) – only the BkPf is capable of encoding a relative clause in this way. Even the Imperfective, which is otherwise morphologically similar to the BkPf, makes the presence of an overt relativizer obligatory: compare sentence (2) with its counterpart (16).

This analysis in turn helps us understand the structure of (75b), here repeated:

- (75) b. LTG Ne keka tekēle kemēm ve gil o si.  
 ART yam some 1EXCL:PL BKPF<sub>1</sub> dig out BKPF<sub>2</sub>  
 '(These are) a few yams we have dug out.'

An initial approach could have proposed analysing (75b) as consisting of a single clause, with a single predicate (*ve gil o si*). In this case, the unusual sentence-initial position of the object noun phrase (*ne keka tekēle*) would probably be explained as a form of left-dislocation. However, this analysis does not hold, for two reasons: formally, the whole sentence is uttered under a single phrase contour with no pause, which makes it incompatible with a topic-focus pattern; and semantically, the function of the initial NP is not that of a topic (\**These yams...*), but of a predicate (*These are some yams...*). This sentence can only be properly analyzed if one remembers that the Torres languages do not make use of copula for noun predicates, i.e. nouns and noun phrases are directly predicative [see §2.1.3, ex.(6)]. Consequently, an appropriate syntactic analysis for (75b) would posit not one clause, but two: first, the whole sentence consists of a zero subject followed by its NP predicate: [*These are*] {*a few yams we have dug out*}; second, the clause *we have dug out* constitutes a relative clause (marked by the BkPf) that is embedded within that main predicate phrase.

Relative clauses marked by the BkPf alone have all the syntactic properties of relative clauses in these languages. They can be embedded within a noun phrase, a prepositional phrase, etc. As mentioned above, the antecedent of the relative can play any syntactic role both in the main clause and in the relative clause itself; and it may also be referred to by a resumptive, anaphoric morpheme within the relative clause (e.g. ē 'there, from it'):

- (87) LTG Ne gerite ni ñwule wulē vete  
 ART octopus AOR:3SG return again place  
 { hōr v' öla t' ē }.  
 3DU BKPF<sub>1</sub> take:TR BKPF<sub>2</sub> OBL:ADV  
 'The octopus went back to the place { they had caught it from }.'

The use of the BkPf in relative clauses is so widespread, that one often hears quite complex sentences such as (88), which superficially consist in a string of juxtaposed clauses, with no obvious indication of their syntactic structure.

- (88) LTG MOWE nie ve velag wahe vin, ni vēn wahe  
 time 3SG IPFV run until up AOR:3SG go until  
 vēn li lylie { nihe ve toge si viēne },  
 thither LOC cave 3PL BKPF<sub>1</sub> stay BKPF<sub>2</sub> underneath  
 { remē mē v' in si viēne }, { ne ñwiē  
 mother his BKPF<sub>1</sub> lie BKPF<sub>2</sub> underneath ART devil  
 ve liē nie t' ē }, nie ni gerage.  
 BKPF<sub>1</sub> replace 3SG BKPF<sub>2</sub> OBL:ADV 3SG AOR:3SG climb  
 'And as he ran all the way up, he managed to reach the cave  
 { (WHERE) they had been staying }, { (WHERE) his mother had been lying },  
 { (AND WHERE) the devil had taken her place }, and he climbed it.'

Apart from the first clause, introduced here by the noun-conjunction *mowe* ‘time, moment’ (§2.2.4), the five remaining clauses lack any subordinator properly speaking. However, the status of the three medial clauses (in braces) as restrictive relative clauses is unambiguous: this is indicated by the Background Perfect, as well as by the presence of locative adverbials (*viēne* ‘underneath’, *ē* ‘there’) whose function is to indicate the syntactic role of their antecedent (the noun *lilie* ‘cave’) within each embedded clause. Ultimately, among the six clauses in (88), only two have the status of informatively new, syntactically main clauses: these are the two Aorist clauses *ni vēn wahe* ‘he reached’ and *ni gerage* ‘he climbed’.

#### 4.2.2.2 Focusing structures

The coding of contrastive focus, in the Torres languages as well as in other languages of north Vanuatu, resorts to a cleft-sentence strategy which is derived from its relativization patterns.

##### 4.2.2.2.1 Contrastive subject focus

The focal constituent, generally a noun phrase, occurs preferably to the left of the sentence – whether via left-dislocation or not – and is immediately followed by a relative clause pointing to the presupposed segment of the utterance.

- (89) Hiw TEKÑWA TAMESō { *pe* ve vegevage vati  
people old REL BKPF<sub>1</sub> talk show  
kema ti ie }.  
1EXCL:PL BKPF<sub>2</sub> OBL:ADV  
(‘It is) the older generation { WHO taught all these stories to us }.’

Clearly, the best way to analyse (89) would be to identify two distinct predicates here, similarly to the analysis of (75b) above. The predicate phrase *vegevage vati* – itself a verb serialization, see (26) – is marked as syntactically dependent as much by the Background Perfect, as by the relativizer *pe*. It is subordinate to the sentence’s main predicate – that is, the nominal predicate *tekn̄wa tamesō* ‘(it is) the elders’.

The syntactic organization of such structures is also reflected in their prosody. A sentence like (89) is uttered with a contrastive accent on the last stressed syllable of the group *tekn̄wa tamesō*. It is followed by a distinctive fall in pitch and intensity on the remainder of the sentence, which is typical of presupposed elements in cleft-constructions:

[təkn̄ʷa ,tamə'so ↓<sub>pə bə ßəyəßəyə bəti kəma ti 'iə]</sub>

The analysis of (89) may also apply to a slightly different form of focusing pattern, one that lacks any formal relativizer. Consider (90):

- (90) Hiw TEKÑWA TE TOGE ve ūak ne gengon ti.  
people from Toga BKPF<sub>1</sub> make ART meal BKPF<sub>2</sub>  
[lit. THE TOGA PEOPLE<sub>[focus]</sub> { made the feast }<sub>[background]</sub>]  
(‘It was) the Toga people (who) organized the feast.’

A first glance at a sentence like (90), which consists of the sequence NP+VP, might have suggested that we are simply dealing with the syntax of a single sentence, with a subject followed by its predicate. However, following the reasoning above for (89), this sentence (90) can rather be shown to consist of two syntactically hierarchized clauses.

The predicate phrase *ve... ti*, which is pragmatically presupposed in the context, would thus be a relative clause with no relativizer, as in (75b) above. The phrase *tekn̄wa te Toge*, to which this relative clause attaches, is pragmatically the focus of the sentence, and syntactically its matrix (NP) predicate. In other words, the syntactic structure of a focusing sentence like (90) is once again parallel to the NP predicate (75b) above:

- (91) NOUN PHRASE + VERB PHRASE with BKPF  
= { nominal equational clause<sub>1</sub> + relative clause<sub>2</sub> (without relativizer) }

The difference between the simple relative clause of (75b) and the focusing structure (90) lies essentially in the prosody. Thus, (90) contrasts a stressed segment with an unstressed one, exactly like (89) above:

[təkn̄ʷa tə 'tɔyə ↓<sub>pə ⁹Lak nə y ən'yan ti]</sub>

##### 4.2.2.2.2 Biclausality and the negation test

The biclausal analysis under (91) is confirmed by certain syntactic tests, such as negation. In principle, the negator is a member of the TAM paradigm (§2.1.2), which means that it normally occurs in the same slot as the corresponding affirmative TAM marker, on the initial boundary of the negated predicate phrase. For example, a standard Perfect like (92a) would be negated as in (92b):

- (92) a. Hiw Tekñwa te Toge nē ūak ne gengon ti.  
people from Toga PRF<sub>1</sub> make ART meal PRF<sub>2</sub>  
‘The Toga people organized a feast.’  
b. Hiw Tekñwa te Toge tati ūak ne gengon.  
people from Toga NEG:REAL make ART meal  
[ordinary negation, no contrastive focus]  
‘The Toga people didn’t organize a feast.’ → 1 CLAUSE

But the sentence’s overall structure turns out to be different when the negation affects a Background Perfect sentence such as (90). Instead of combining with the verb *ūak* as in

(92b), the negator then affects the initial noun phrase of the sentence, thereby proving it has the syntactic status of a predicate:

- (93) Hiw *Tati* *tek̄iwa* *te* *Toge* *ve* *řak* *ne* *gengon* *ti.*  
 NEG:REAL people from Toga BKPF<sub>1</sub> make ART meal BKPF<sub>2</sub>  
 [negation of contrastive focus pattern]  
 '{ It's not the Toga people }<sub>[focus]</sub> (who) organized the feast<sub>[bkg]</sub>'  
 → 2 CLAUSES

In sum, (90) consists not just of a subject phrase with its predicate, but of two predicates: it must be analyzed as a genuine cleft construction.

Finally, exactly the same analysis could be conducted to account for Example (79), mentioned in §4.1.4 and repeated below:

- (79) Hiw *Noke* *ve* *tot* *ti.*  
 1SG BKPF<sub>1</sub> carve BKPF<sub>2</sub>  
 [lit. '{ (it's) I }<sub>[focus]</sub> (who) { carved it }<sub>[background]</sub>']  
 'I made it!'

While the shortness and simplicity of (79) would spontaneously suggest we are dealing with a monoclausal SV(O) sentence just like its English translation, it turns out that a more accurate analysis would have to parse it into two distinct clauses: a direct noun predicate (*noke*)<sup>31</sup> followed by a relative clause with no relativizer (*ve tot ti*).<sup>32</sup> Thus the negation of (79) would be parallel to (93) above:

- (79') Hiw *Tati* *noke* *ve* *tot* *ti.* *Temo-k.*  
 NEG:REAL 1SG BKPF<sub>1</sub> carve BKPF<sub>2</sub> father-1SG  
 '{ (It's) not I }<sub>[focus]</sub> { (who) carved it }<sub>[bkg]</sub>. (It's) my father.'

#### 4.2.2.2.3 Contrastive focus of non-subjects

The analysis proposed above for the contrastive focus of subject noun phrases can be extended to other syntactic functions and other parts of speech. Indeed, we know (from §2.1.3) that the ability to constitute a direct predicate – with no copula – is not

31. Ex. (8) above illustrates the same pronoun *noke* 'it's' me' in a direct NP predicate structure.

32. Evans (2007), in his article on "insubordination", cites similar instances of 'hidden' cleft constructions in certain Australian languages. For example, the language Ngandi (Evans 2007: 414, after Heath 1985) expresses subject focus by combining an ordinary subject NP with a verb form that is formally marked as subordinate (with *ga*): e.g. *ni-deremu ni-GA-rudu-yi*, literally '[it's] the man [who] went<sub>subord</sub>'. The structural similarity with our proposed analysis (91) is worthy of notice here: in both cases, the surface form of the sentence seems to consist of a single clause, where underlyingly there are two.

only characteristic of nouns and noun phrases, but in fact of most other parts of speech and syntactic constituents.

It is thus possible to interpret all focus constructions as *biclausal* sentences, along the lines of (91). The focus phrase forming a direct predicate may be e.g. an adverb (94) or a predicative demonstrative (95):

- (94) Hiw *Ve* *řak* *ti* *ñWĒNA?*  
 BKPF<sub>1</sub> make BKPF<sub>2</sub> how  
 [lit. { made it }<sub>[background]</sub> HOW<sub>[focus]</sub>?]  
 'How was it made?'

- (95) LTG *Noke* *ve* *vēn* *ve* *tun* *si* *Vave* *PE* *NOK!*  
 1SG BKPF<sub>1</sub> go BKPF<sub>1</sub> buy BKPF<sub>2</sub> Vava FOC this  
 [lit. { I went to buy on Vava }<sub>[background]</sub> { (it's) THIS }<sub>[focus]</sub>]  
 'THIS is what I bought on Vava island.'

In those cases too, the BkPf clause can be analyzed as a relative clause followed by its matrix predicate.

The case for this biclausal analysis is even stronger when the asserted phrase is fronted, as commonly happens in cleft focus constructions. As mentioned in §2.1.1, the constituent order is normally SVO. When the asserted element coincided with the subject of the backgrounded verb, as in (90) or (79) above, the focus construction involved no displacing of the phrase under focus; its pragmatic status was only indicated by the prosody (and of course, indirectly, by the BkPf in the rest of the sentence). But when fronting affects an object or another complement whose normal position is after the predicate, then the disrupted syntax of the sentence makes it clear that we are dealing with a biclausal structure.

For example, compare the non-contrastive sentence (96a) – with standard word order and the regular Perfect – and its contrastive counterpart (96b):

- (96) a. LTG *Gide* *na* *vēn* *si* *me* *ē* *ne* *mesale* *pek.*  
 1INCL:PL PRF<sub>1</sub> go PRF<sub>2</sub> hither OBL:PREP ART road this  
 'We came through this road.'  
 → 1 CLAUSE  
 b. LTG *NE* *MESALE* *PEK* *gide* *ve* *vēn* *si* *me* *ē.*  
 ART road this 1INCL:PL BKPF<sub>1</sub> go BKPF<sub>2</sub> hither OBL:ADV  
 [lit. '(it is) THIS ROAD (that) we came through (it)?']  
 'THIS is the road we came through.'  
 → 2 CLAUSES

(96b) shows fronting of the focal element, in the form of a predicate noun phrase (*ne mesale pek* 'it is] this road'). The remainder of the sentence, which is marked as BkPf, has the syntactic status of a relative clause. Specifically, the antecedent *mesale* 'road' is anaphorically indexed by the locative preposition-adverb *ē* ('there, through it') – in accordance with the typical syntax of relative clauses, as in (87) above. The resulting

double-zero relative clause – i.e. zero relativizer, zero anaphora on the preposition – happens to be structurally close to its English equivalent: *(it is) THIS ROAD {Ø we came through Ø}*.

We saw earlier that the surface form of subject-focusing sentences like (79) shows some structural ambiguity, to the point that certain tests were required to determine their underlying syntax (§4.2.2.2). This is not necessary with other contrastive focus cleft constructions such as (96b), because they are transparent in this regard.

In sum, a predicate marked as Background Perfect must always be understood as forming a subordinate clause – even when superficially it may seem to form the sole verb of the utterance. The pragmatic center of assertion, as well as the syntactic center of the sentence, are to be sought outside of its boundaries.

#### 4.2.2.3 Wh-questions and the Background Perfect

Finally, a contrast similar to (96a–b) can be found in the structure of questions. At first sight, the different choice of aspect between (97a) and (97b) is difficult to explain:

- (97) a. LTG *Nike na vegevage si mi paie?* → (?ve vegevage si...)  
           2SG PRF<sub>1</sub> talk PRF<sub>2</sub> with who  
           [lit. You were talking to whom?]  
           ‘Who were you talking to?’ [STANDARD PERFECT]  
       b. LTG *Paie ve vegevage si mē-ke?* → (\*na vegevage si...)  
           who BKPF<sub>1</sub> talk BKPF<sub>2</sub> with-you  
           ‘Who was talking to you?’ [BACKGROUND PERFECT]

The rule that is empirically observed, and illustrated by (97a–b), is given in (98):

- (98) In content questions referring to a completed event (perfect), the verb will normally take the REGULAR PERFECT if the question word comes after the verb; but it must be marked as BACKGROUND PERFECT if the question word precedes the verb (whether by *wh*-movement or not).

The explanation for this unexpected asymmetry has to do with the placement of sentential focus, which in content questions systematically falls upon – or includes – the question word. In (97a), which is unmarked for word order, the sentence-final position of the question word *paie* is compatible with the interpretation of the whole predicate (including its complement) as falling under the pragmatic focus of the utterance. In (97b) however, the sentence-initial position of *paie* attracts stress and sentential focus, yielding a sentence shape that is strongly reminiscent of focalising structures such as (79) or (96b). A consequence of this sentence-initial focus is that the rest of the sentence has to be coded as informationally defocused, which explains the use of the Background Perfect here. Once again, the most appropriate analysis of (97b) is to consider it as biclausal, similarly to (91) above. In other words, what we have here is literally:

- (97) b. ‘{ (it is) WHO }<sub>[focus]</sub> (the one that) { was talking to you }<sub>[background]</sub>?’

Such a formal TAM contrast between (97a–b), depending on the placement of the question word, is unique to the Torres languages, and unknown elsewhere in the region. Furthermore, it is even quite particular within these two languages, as it is restricted to questions whose verbal aspect is a perfect. Uncommon though it may be, this contrast can however be explained by the internal logics of these languages, in terms of the handling of informational hierarchy and predicate dependencies.

#### 4.3 The Background Perfect: Summary

The various patterns characteristic of the Background Perfect are summarized in Table 5.

Table 5. The close links between the Background Perfect and clause dependency: A summary

SYNTAX	FUNCTIONAL VALUE	EXAMPLES
no subordination	clause topicalization & backgrounding	(77b)
combines with subordinators	realis background (restrictive) relative clauses	(82)–(83)
directly encodes subordination	realis background clause in cleft focus patterns	(89)
	realis background (restrictive) relative clauses	(85)–(88)
	realis background clause in cleft focus patterns	(90)–(96b)
	⇒ question sentences if <i>wh</i> -word is fronted	(97b)

#### 5. Conclusion

Hiw and Lo-Toga, the two languages of the Torres islands, possess a wealth of formal devices for encoding clause dependency, and make regular use of them with most of their TAM markers. However, this paper has shown that two TAM categories – the Subjunctive and the Background Perfect – present different behaviour when it comes to handling interclausal relations. While they are both compatible with regular subordinators, they also show a marked tendency to do without them, and to be used alone as a subordinating strategy in its own right.

Obviously, the two cases under study differ in many respects, if only because they do not come under the same discourse constraints:

- the Subjunctive contrasts with other irrealis markers, in lacking the necessary information on the clause’s MODALITY STATUS and ILLOCUTIONARY FORCE.
- the Background Perfect contrasts with other realis categories (especially with the regular Perfect), in marking its target predicate as PRAGMATICALLY PRESUPPOSED.

One characteristic that is nevertheless shared by these two components is that they both affect the pragmatic well-formedness of an utterance. A sentence, if irrealis, needs to have some form of illocutionary force; and likewise, an utterance must include at least some new, asserted segment. In my interpretation, the absence of either of these two elements in a clause is precisely what makes it unable to form a sentence on its own, and makes it dependent, both functionally and syntactically, upon external predicates and clauses.

In sum, different as they may be, these two patterns essentially obey the same underlying mechanism, which justifies their comparison. In both cases, the key to the syntactic structures attested is a form of pragmatic indeterminacy, or *pragmatic demotion*, that is inherently conveyed by the TAM marker.

The two patterns illustrated in this paper are specific to Hiw and Lo-Toga, and make these two languages original, even in comparison with the nearby languages of north Vanuatu. Yet they also show a form of universal relevance. They remind us that the existence of formal, dedicated subordinators is not the sole key to the syntax of interclausal relations; and that patterns of clause dependency can also result, albeit indirectly, from a clause's pragmatic properties and semantic profile. This is another illustration of how the formal structures of languages are regularly shaped and renewed through the functional constraints that weigh upon communication.

## Abbreviations

Examples are glossed according to the Leipzig rules. More specific abbreviations are listed below.

AFF	affirmative	LTG	Lo-Toga
AOR	Aorist	M	masculine
APPL	applicative	NEG:EXIST	Negative existential
ART	article	OBL	oblique
BKPF	Background Perfect	POC	Proto Oceanic
CAUS	causative	POSS	possessive marker
COMP	complementizer	POT	Potential
CPLT	Complete aspect	PRF	Perfect
CTFC	Counterfactual	PROSP	Prospective
DU	dual	QUOT	quotative
FOC	focus marker	REL	relativizer
FUT	Future	RESULT	resultative
IPFV	Imperfective	S	subject clitic
IRR	irrealis	SBJV	Subjunctive
ITER	iterative	STAT	Stative
HIW	Hiw	TR	transitive verb
LOC	locative marker		

## References

- Aikhenvald, Alexandra Y. 2005. Serial verb constructions in a typological perspective. In *Serial Verb Constructions: A Cross-linguistic Typology* [Explorations in Linguistic Typology], Robert M.W. Dixon & Alexandra Y. Aikhenvald (eds), 1–68. Oxford: OUP.
- Bril, Isabelle. 2004. Complex nuclei in Oceanic languages: Contribution to an areal typology. In *Complex Predicates in Oceanic Languages: Studies in the Dynamics of Binding and Boundedness*, Isabelle Bril & Françoise Ozanne-Rivierre (eds), 1–48. Berlin: Mouton de Gruyter.
- Bril, Isabelle & Ozanne-Rivierre, Françoise (eds). 2004. *Complex Predicates in Oceanic Languages: Studies in the Dynamics of Binding and Boundedness*. Berlin: Mouton de Gruyter.
- Bybee, Joan, Perkins, Revere & Pagliuca, William. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago IL: University of Chicago Press.
- Chappell, Hilary. 2008. Variation in the grammaticalization of complementizers from *verba dicendi* in Sinitic languages. *Linguistic Typology* 12(1): 45–98.
- Comrie, Bernard. 1976. *Aspect. An Introduction to the Study of Verbal Aspects and Related Problems*. Cambridge: CUP.
- Cristofaro, Sonia. 1998. Deranking and balancing in different subordination relations: A typological study. *Sprachtypologie und Universalienforschung* 51: 3–42.
- Cristofaro, Sonia. 2003. *Subordination* [Studies in Typology and Linguistic Theory]. Oxford: OUP.
- Cristofaro, Sonia. 2008. Purpose clauses. In *The World Atlas of Language Structures Online*, Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds), Ch. 125. Oxford: OUP.
- Crowley, Terry. 1987. Serial verbs in Paamee. *Studies in Language* 11: 35–84.
- Crowley, Terry. 2002. *Serial Verbs in Oceanic: A Descriptive Typology*. Oxford: OUP.
- Crowley, Terry. 2004. *Bislama Reference Grammar*. Honolulu HI: University of Hawaii Press.
- Dik, Simon. 1989. *The Theory of Functional Grammar*, Part 1: *The Structure of the Clause* [Functional Grammar Series 9]. Dordrecht: Foris.
- Durie, Mark. 1997. Grammatical structures in verb serialization. In *Complex Predicates*. Alex Alsina, Joan Bresnan & Peter Sells (eds), 289–354. Stanford: CSLI.
- Ernout, Alfred & Thomas, François. 1953[1993]. *Syntaxe latine*. Paris: Klincksieck.
- Evans, Nicholas. 2007. Insubordination and its uses. In *Finiteness. Theoretical and Empirical Foundations*, Irina Nikolaeva (ed.) 366–431. Oxford: OUP.
- Foley, William A. & Olson, Mike. 1985. Clausehood and verb serialization. In *Grammar inside and outside the clause. Some approaches to theory from the field*. Johanna Nichols & Anthony C. Woodbury (eds), 17–60. Cambridge: CUP.
- François, Alexandre. 1997. La subordination sans marques segmentales: Formes de dépendance interpropositionnelle dans le discours. Mémoire de DEA. MA thesis, Université Paris-III Sorbonne Nouvelle.
- François, Alexandre. 2003. *La sémantique du prédicat en mwotlap (Vanuatu)*. Collection Linguistique de la Société de Linguistique de Paris, 84. Louvain: Peeters.
- François, Alexandre. 2004. Chains of freedom: Constraints and creativity in the macro-verb strategies of Mwotlap. In Bril & Ozanne-Rivierre (eds), 107–143.
- François, Alexandre. 2005a. Diversité des prédicts non verbaux dans quelques langues océaniennes. In *Les constituants prédictifs et la diversité des langues: Actes de la Journée de la Société de Linguistique de Paris* [Mémoires de la Société de Linguistique de Paris], Jacques François & Irmtraud Behr (eds), 179–197. Louvain: Peeters.

- François, Alexandre. 2005b. Unraveling the history of the vowels of seventeen northern Vanuatu languages. *Oceanic Linguistics* 44(2): 443–504.
- François, Alexandre. 2006. Serial verb constructions in Mwotlap. In *Serial Verb Constructions: A Cross-linguistic Typology* [Explorations in Linguistic Typology], Robert M.W. Dixon & Alexandra Y. Aikhenvald (eds), 223–238. Oxford: OUP.
- François, Alexandre. In press. Verbal aspect and personal pronouns: The history of aorist markers in north Vanuatu. In *A Festschrift for Robert Blust*, Alexander Adelaar & Andrew Pawley (eds). Canberra: Australian National University.
- François, Alexandre. In prep. From deictics to clause linkers. Discourse deixis, topicalization and clause backgrounding strategies in the languages of the Banks islands (Vanuatu).
- Givón, Talmy. 1984/1990. *Syntax. A Functional-typological Introduction*. Amsterdam: John Benjamins.
- Haspelmath, Martin. 2007. Pre-established categories don't exist – consequences for language description and typology. *Linguistic Typology* 11(1): 119–132.
- Haspelmath, Martin, Dryer, Matthew, Gil, David & Comrie, Bernard. (eds). 2008. *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. <<http://wals.info/>>.
- Heath, Jeffrey. 1985. Discourse in the field: clause structure in Ngandi. In *Grammar inside and outside the Clause: Some Approaches to Theory from the Field*, Johanna Nichols & Anthony C. Woodbury (eds), 89–110. Cambridge: CUP.
- Heine, Bernd & Kuteva, Tania. 2002. *World Lexicon of Grammaticalization*. Cambridge: CUP.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representation of Discourse Referents* [Cambridge Studies in Linguistics 71]. Cambridge: CUP.
- Launey, Michel. 1994. *Une grammaire omniprédictive: Essai sur la morphosyntaxe du nahuatl classique*. Sciences du Langage. Paris: CNRS.
- Lemaréchal, Alain. 1989. *Les parties du discours. Syntaxe et sémantique* [Linguistique Nouvelle]. Paris: Presses Universitaires de France.
- Lemaréchal, Alain. 1992. Extension possible de la notion d'orientation aux subordonnées complétives et leurs équivalents. *Bulletin de la Société de Linguistique de Paris* 87 (1): 1–35.
- Mühlhäusler, Peter, Dutton, Thomas E. & Romaine, Suzanne. 2003. *Tok Pisin Texts: From the Beginning to the Present* [Varieties of English around the World T9]. Amsterdam: John Benjamins.
- Noonan, Michael. 1985. Complementation. In *Language Typology and Syntactic Description*, Timothy Shopen (ed.), 42–140. Cambridge: CUP.
- Shopen, Timothy. (ed.) 1985. *Language Typology and Syntactic Description*, Vol.2. Cambridge: CUP.
- Stassen, Leon. 2000. AND-Languages and Wirth-Languages. *Linguistic Typology* 4: 1–54.
- Thompson, Sandra A. & Longacre, Robert E. 1985. Adverbial clauses. In *Language Typology and Syntactic Description*. Timothy Shopen (ed.), 169–234. Cambridge: CUP.
- Tomlin, Russell. 1985. Foreground-background information and the syntax of subordination. *Text* 5: 85–122.
- van der Auwera, Johan & Plungian, Vladimir. 1998. Modality's semantic map. *Linguistic Typology* 2: 79–124.
- van der Auwera, Johan, Dobrushina, Nina & Goussev, Valentin. 2008. Imperative-hortative systems. In Martin Haspelmath, Matthew Dryer, David Gil & Bernard Comrie (eds), Ch. 72. <<http://wals.info/feature/description/72>> 15 November, 2008.

# Lexical and Structural Etymology

---

Edited by  
Robert Mailhammer

DE GRUYTER  
MOUTON

Alexandre François

**Shadows of bygone lives: The histories of spiritual words in northern Vanuatu**

## 1 Studying etymology with unwritten languages

Many linguists are familiar with etymologies in Indo-European languages, and how they help us discover invisible threads between words that have become widely separated in form or in meaning.<sup>1</sup> Etymology teaches us that *miracle*, *marvel* and *mirror* all stem from a single Latin verb *mīrāri* ‘wonder, admire’ – respectively via its derived nouns *mīrāculum*, Late Latin *mīrābilis*, and Old French *mireor*. The deep-reaching insights of Proto Indo-European reconstruction even allow us to find connections between the words *idea*, *view*, *advice*, *story* and *witty*: they all stem ultimately from a single PIE root whose form can be reconstructed as \**weyd* ‘see, know’ (cf. Chantraine 1968: 796). Similar etymological links can be drawn between *quick*, *vital*, *bio* and *zoo* <\**gʷyeh*₃ ‘live’ (Chantraine 1968: 176); between *culture*, *wheel*, *colonial*, *polar*, *cyclone*, *bike* and *chakra* <\**kʷel* ‘turn, make a turn’ (Chantraine 1968: 878); or between *diary*, *Zeus*, *deity*, *journey*, *diva* and *Tuesday* <\**dey-* ‘bright, shine’ (Chantraine 1968: 399).

A fair proportion of our etymological knowledge about Indo-European languages rests on the careful philological analysis of written documents (e.g. Benveniste 1969), some of which are extremely ancient. This is so true that the existence of a long tradition of writing might be taken by some as a preliminary condition for any serious endeavour in the realm of etymology. Could the same sort of research be carried out on languages with no written tradition at all, and for which we can only access present-day forms? The answer to this question, of course, is positive. While *philology* is restricted to the study of written documents from the past, the study of *etymology* is a branch of historical linguistics, which can also take unwritten languages as its object. Simply, instead of involving a comparison between various historical attested stages of a single language, etymology will rest here on comparison *across* related languages.

---

<sup>1</sup> I am grateful to Andrew Pawley, Malcolm Ross, Robert Blust, Sabine Hess, Margaret Jolly, Siva Kalyan and Maïa Ponsonnet, for their helpful comments on earlier versions of this article. I also thank Marie Durand who supplemented my Mwerlap data with new words and cultural comments. Finally, such a study would not have been possible without the help and kindness of the many people from northern Vanuatu who patiently taught me their languages.

As other chapters in this volume show, etymological research based on previously unwritten languages is not only feasible, but it has long been carried out in various parts of the world. Oceanic languages, which are the focus of the present chapter, have been the object of comparative research as early as the 18th century, an effort which culminated in Dempwolff (1938) demonstrating the unity of the Oceanic subgroup. Etymological research in this domain is continued nowadays by various scholars – with a special mention for the *Oceanic lexicon project* which Malcolm Ross, Andrew Pawley and Meredith Osmond have led since 1998 (see references). These Oceanic studies form part of a larger domain of research on Austronesian etymologies, which for several decades has been dominated by the prolific work of Robert Blust.

Rather than undertaking a general survey of the etymological research undertaken on these vast language families, the present study will focus on a set of 17 closely related Oceanic languages spoken in northern Vanuatu. I will narrow the domain of investigation to the “spiritual” field: words for *mind*, *soul*, *spirit*, *ghost*, *shaman*, *supernatural powers* and other related concepts. This survey will lead me to highlight the essential role played by ancestral spirits in the cultural landscape and social representations of northern Vanuatu communities. Spirits are present in the land, in rocks and magic stones, in the depths of caves and volcanoes. They live on in the secret societies where their presence is symbolically reenacted. Ancestral spirits are the source of various dances and musical instruments, of the language of poetry, of many myths and stories; their magic power can be harnessed by dancers, poets, shamans, sorcerers. All these connections across different social domains have left traces in the modern languages – yet these traces are sometimes hidden, and can only be detected by the tools of the historical linguist.

In a way, the present study can be read as an introduction to the spiritual world of island Melanesia, from the particular angle of etymology, or “linguistic archaeology”. It should be of interest to linguists, but also to anthropologists and other scholars in social sciences, with no specific background in linguistics. Even though the phonetic transcriptions of word forms will necessarily comply, for the sake of accuracy, with the sometimes opaque conventions of the *International Phonetic Alphabet* (IPA), the non-linguist reader should still be able to follow the reasonings about semantic description and reconstruction.

My aim will be to illustrate a point that historical linguists have already shown many times (see Biggs, Blust, Pawley, Ross in the references): namely, that lexical comparison among unwritten languages can provide valuable insights in the domain of ethnographic analysis and culture history. For example, I will show that the word *nɔ-tɔm* in the language Mwerlap, which refers to ‘a magical sign put up to keep ghosts away’, is cognate with *ne-tekp<sup>w</sup>* ‘graveyard’

in the neighbouring language Mwotlap (§4.5.2); and that both words ultimately reflect a Proto Oceanic root \*ta<sup>m</sup>bu ‘off limits, forbidden’ – itself the source of English *taboo*. In proposing such hypotheses, scholars need to constantly bear in mind the risk of proposing false etymologies based on mere similarity or impressionistic likelihood. This is where the power of the comparative method becomes crucial, as it defines a solid methodology based on the observation of regular sound correspondences, as well as the identification of convincing semantic links.

This study will unfold in three major parts. After locating the target languages in their geographic and historical context (Section 2), I will outline the methodological principles that underlie the search for reliable etymologies (Section 3). Once the general points of the method are defined, section 4 will delve into the fine-grained detail of spiritual words in this part of Melanesia, as we reconstruct the semantic evolution of about 80 roots or cognate sets. This survey will hopefully show how the discipline of etymology, as long as it complies with strict linguistic principles, is able to uncover semantic networks of the past – even when centuries of phonetic or semantic change have made them difficult to detect in modern languages.

## 2 The Torres–Banks languages of north Vanuatu

### 2.1 The languages of Vanuatu

One of the archipelagoes of Island Melanesia, the modern country of Vanuatu – known as the New Hebrides until its independence in 1980 – hosts the world’s highest density of languages *per capita* (Crowley 2000). A total of 106 vernacular languages have been recorded there (Tryon 1976; Lynch and Crowley 2001) for a current population of just 234,000 inhabitants. Most of that population still lives in rural areas, where they essentially pursue traditional activities of farming and fishing. Their current way of life largely continues that of their ancestors, even though modern times have brought in various forms of cultural change – such as the Christian religion, or cash economy. Figure 1 situates Vanuatu within Island Melanesia and the Pacific.

The present study focuses on the languages of the Torres and Banks Islands, in the northernmost area of Vanuatu. Figure 2 provides a linguistic map of the area. Each language name is followed by an approximate number of speakers, and by a three-letter abbreviation which will be used throughout this paper.

Until recently, the only detailed source of documentation about these northern languages was the pioneering work of the Anglican missionary and linguist Robert

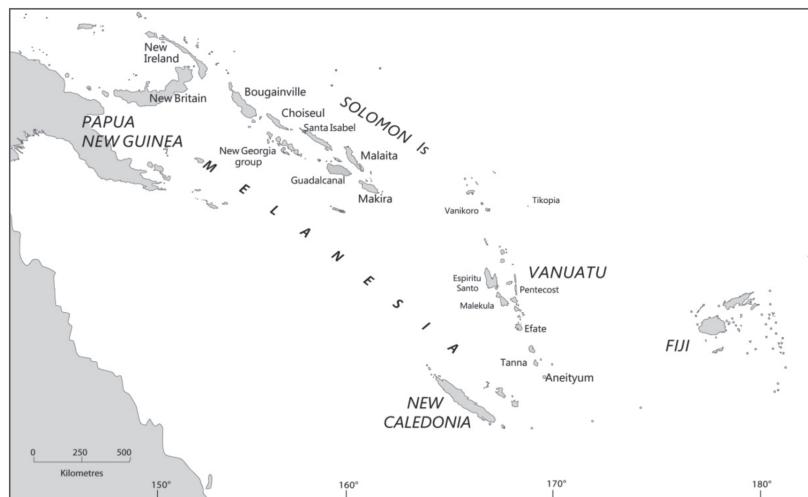


Figure 1: Situation of Vanuatu in the Pacific

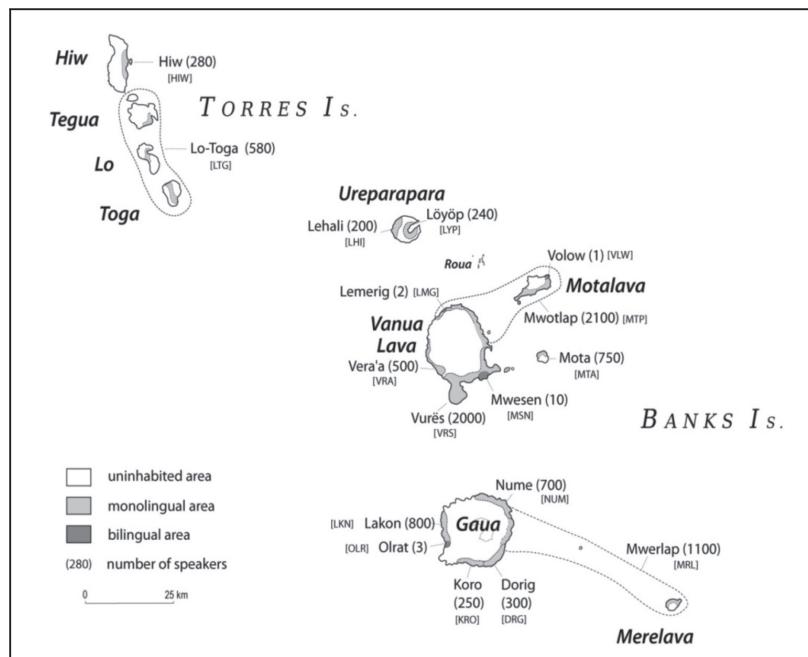


Figure 2: The 17 languages of northern Vanuatu (Torres and Banks Is.)

Codrington, at the end of the 19th century (see references). His grammatical and ethnographic descriptions were mostly focused on the small island of Mota, yet occasionally reached out to neighbouring islands. While anthropologists have maintained a constant interest in the region (Ivens 1931; Vienne 1972, 1984; Lanouguère-Bruneau 2002; Mondragón 2004; Hess 2009), no linguistic study other than wordlists (Tryon 1976) was published about the Torres–Banks area during the whole of the 20th century. The present chapter rests on firsthand data I collected during a number of field trips in Vanuatu since 1997.<sup>2</sup>

## 2.2 Three millennia of diversification

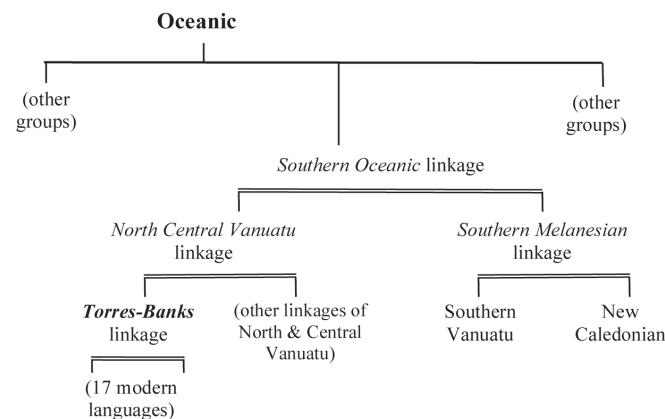
The linguistic density of Vanuatu does not reflect deep genetic diversity: all these languages share the same ancestor, Proto Oceanic – itself a member of the vast Austronesian family. There are strong arguments for equating Proto Oceanic with the language that was spoken by the bearers of the archaeological culture known as “Lapita” – the seafaring population who first settled the archipelagoes between the Solomons and Western Polynesia about 3000 years ago (Pawley 2003, 2007, 2008, 2010). In a way similar to the fragmentation of Latin into a multitude of Romance languages and dialects, the linguistic diversity observed today in Vanuatu results from three millennia of *in situ* diversification from what was once a single language spoken across a vast social network (François 2011a).

During the centuries following its initial settlement, Vanuatu formed a vast dialect continuum in which communalects remained in constant contact through trade, interisland marriage and other forms of alliances. Every time a linguistic innovation emerged somewhere in the network, it would diffuse to a more or less extended portion of the network. The isogloss it defined was sometimes limited to just a village, sometimes to several islands, and sometimes swept through even larger territories as it expanded across entire archipelagoes (François 2011a,b). Rather than yielding neat subgroups, this wave-like process of diversification naturally resulted in a map of constantly intersecting isoglosses. The modern outcome is a vast web of “linkages,” i.e. linguistic areas resulting from the breakup of an earlier dialect continuum (Ross 1988: 8).

Vanuatu as a whole can be considered a linkage, which in turn includes various loosely bound linkages of different scales. This chapter will especially

<sup>2</sup> I occasionally compared my fieldwork data with Codrington and Palmer (1896) for Mota, and with Hyslop (n.d. *a–b*) for Vurès and Vera'a. Marie Durand has provided some personal comments regarding Mwerlap.

mention “North Central Vanuatu” – whose putative protolanguage PNCV is being reconstructed by Clark (2009) – as well as the “Torres–Banks linkage” (François 2011b), the focus of this study. Figure 3 situates the latter in the context of other subgroups and linkages of Oceanic. Following conventions proposed by Ross (1988), double lines indicate linkages as opposed to discrete subgroups.



**Figure 3:** Proto Oceanic has split into various dialect networks, which have given rise to loosely-bound linkages of various sizes

In the remainder of this paper, lexical reconstructions will correspond to the level of “Proto Torres–Banks” – the common ancestor of the Torres–Banks linkage – yet I will occasionally also cite reconstructions at the higher levels of Proto North-Central Vanuatu (PNCV, Clark 2009) and Proto Oceanic (POc).

### 2.3 Diversification and cognate identification

Even though they remained in mutual contact during the three millennia of their settlement, the languages of northern Vanuatu have diverged so as to lose mutual intelligibility.

As an illustration of the modern diversity, (1) gives the term for ‘woman’ in the 17 languages of the area. Languages are ranked geographically from northwest to southeast, and designated by their abbreviation (see Figure 2):<sup>3</sup>

<sup>3</sup> All forms are given in IPA. Nouns are given in their citation form, which occasionally includes a prefix of the form *n-* or *nV-*: e.g. *na-lkpʷəβin*. This is a noun article which many languages have incorporated into their phonological word (François 2007). Even though this article can be reconstructed (as \**na*), I will leave it out of my lexical reconstructions.

- (1) ‘woman’: HIW *jəkʷen*; LTG *lakʷəβinə*; LHI *n-lɔkβen*; LYP *n-lɪkpʷiɛn*; VLW *n-leŋgbʷeβin*; MTP *na-lkpʷəβin*; LMG *n-reŋe*; VRA *reŋe*; VRS *rekpʷe*; MSN *rekpʷe*; MTA *taβne*; NUM *tawa*; DRG *rkpʷa*; KRO *rakpʷa*; OLR *rakpʷa*; LKN *γamtv*; MRL *na-βaβeān*.

The processes whereby northern Vanuatu languages diverged from their common ancestor took mostly two forms: sound change and lexical replacement (François 2011b: 192). A form like Mwerlap *na-βaβeān* is conservative of the Proto Oceanic etymon for ‘woman’, namely \**pine* or \**pa-pine*, and has simply gone through sound change. Mota *taβne* reflects a slightly different etymon \**ta-pine* (§4.1.2). Likewise the first six forms of (1), whose protoform reconstructs as \**lombʷe-βine*, also keep a vestigial trace of POc \**pine* – even though it was occasionally made invisible by subsequent sound change, as in Hiw *jəkʷen*. By contrast with these cases of partial cognacy, the remaining nine forms of (1) illustrate various events of lexical replacement with material that bears no relation at all to the POc etymon.

The amount of phonetic erosion and lexical change that took place in the history of northern Vanuatu languages makes it necessary to define strict methodological principles with the aim to secure historical hypotheses, and minimize the risk of false etymologies.

## 3 The methodology of lexical reconstruction

The principal steps to be taken in lexical reconstruction and etymology can be outlined as follows. First, knowledge of regular sound correspondences across languages makes it possible to ascertain which words are cognate. Solid hypotheses about sound change enable the reconstruction of a protoform for each cognate set, i.e. the most likely phonological form taken by the word that was once at the source of the modern cognate set. Finally, the meanings of modern reflexes form the basis on which earlier meanings, and processes of semantic change, can be hypothesized.

The following pages will briefly illustrate the process with a few examples. This will serve as an introduction to the etymological explorations presented in Section 4.

### 3.1 Establishing regular correspondences

In spite of the occasionally extreme linguistic divergence illustrated in (1), the comparison of lexical material across northern Vanuatu does bring about regular

correspondences. For example, consider the modern words for ‘stingray’ (2) and for ‘hibiscus’ (3):

- (2) ‘stingray’: HIW *βɔ̊gL*; LTG *βer*; LHI *n-βæj*; LYP *n-βij*; VLW *n-βij*; MTP *ne-βej*; LMG *n-βer*; VRA *fer*; VRS *βœr*; MSN *βer*; MTA *βar*; NUM *fer*; DRG *βar*; KRO *βeär*; OLR *βaj*; LKN *βæ:*; MRL *ne-βer*.
- (3) ‘hibiscus’: HIW *βɔ̊gL*; LTG *βor*; LHI *n-βvj*; LYP *n-βij*; VLW *n-βej*; MTP *ne-βej*; LMG *n-βer*; VRA *fer*; VRS *βœr*; MSN *βor*; MTA *βar*; NUM *far*; DRG *βar*; KRO *βeär*; OLR *βaj*; LKN *βa:*; MRL *n-βɔ̊r*.

The correspondences regarding consonants are easy to spot. Thus a word-initial [f] in Vera'a and Nume corresponds to [β] elsewhere. A velar lateral [g<sub>L</sub>] in Hiw corresponds to a trill [r] in some languages, a glide [j] in others, and a long vowel in Lakon (François 2011a: 150). Based on these correspondences, one can formulate hypotheses on the nature of the protophonemes (the source phonemes in the protolanguage) reflected by these modern forms. For example, it is likely that [β] and [f] reflect a former consonant \*β. As for the second set of correspondences, it demonstrably reflects a former apical trill \*r. Such hypotheses on individual segments are a preliminary step before entire words can be reconstructed – as we will see below.

Correspondences with respect to vowels are more complex, yet they also prove highly regular. The forms in (2) ‘stingray’ illustrate a first set of correspondences, outlined in (4):

- (4) Vowel correspondence set I:  
HIW /ɔ:/; LTG /ε:/; LHI /æ:/; LYP /I:/; VLW /I:/; MTP /ε:/; LMG /ε:/; VRA /ε:/; VRS /œ:/;  
MSN /ε:/; MTA /a:/; NUM /ε:/; DRG /a:/; KRO /ɛä:/; OLR /a:/; LKN /æ:/; MRL /ε:/.

A very similar distribution of vowels can be found, for example, in the numeral ‘four’:

- (5) ‘four’: HIW *βɔt*; LTG *βet*; LHI *βæt*; LYP *βet*; VLW *n-βet*; MTP *βet*; LMG *βe?*; VRA *fe?*; VRS *βet*; MSN *βet*; MTA *βat*; NUM *fat*; DRG *βa:t*; KRO *βeä:t*; OLR *βat*; LKN *βæs*; MRL *βet*.

Considering the phonetic diversity of the modern vowels in this correspondence set (4), it may seem impossible to reconstruct the phonetic value of the protovowel it is supposed to stem from. However, the solution can be found if other Oceanic languages are taken into the equation (François 2005). For example, the words for ‘stingray’ and ‘four’ are respectively [fari] and [fati] in

Sungwadia, the language of northern Maewo (Agnès Henri, p.c.). These two forms show a word-final trochaic foot /CáCi/ where Torres-Banks languages have a single syllable CVC: e.g. Sungwadia *fati* corresponds to Lo-Toga *βet*. Many more examples could be cited, showing that the vowel correspondence set in (4) reflects a former trochaic foot \*(C)á(C)i that has undergone a process of umlaut, or metaphony. Even though modern reflexes are sometimes phonetically unexpected – as when Hiw reflects \*(C)á(C)i with a rounded vowel – the processes are highly regular in each language. In other words, (4) above can be rewritten as the result of a historical process of sound change:

- (4') \*(C)á(C)i > HIW /ɔ:/; LTG /ε:/; LHI /æ:/; LYP /I:/; VLW /I:/; MTP /ε:/; LMG /ε:/; VRA /ε:/; VRS /œ:/; MSN /ε:/; MTA /a:/; NUM /ε:/; DRG /a:/; KRO /ɛä:/; OLR /a:/; LKN /æ:/; MRL /ε:/.

A similar reasoning holds for the second correspondence set in ‘hibiscus’ (3), yet with a different set of vowels, as listed in (6):

- (6) Vowel correspondence set II:  
HIW /ɔ:/; LTG /ɔ:/; LHI /v:/; LYP /I:/; VLW /ε:/; MTP /ε:/; LMG /ε:/; VRA /ε:/; VRS /œ:/;  
MSN /ɔ:/; MTA /a:/; NUM /a:/; DRG /a:/; KRO /ɛä:/; OLR /a:/; LKN /a:/; MRL /ɔ̊:/.

If we know that ‘hibiscus’ is *fau* in Samoan and *valu* in Gela of the Solomon Islands (Ross, Pawley, and Osmond 2008: 139), it is safe to conclude that the vowel set (6) reflects a former sequence \*(C)á(C)u, which also went through umlaut, generally with back reflexes rather than front. Thus (6) can be rewritten as a historical process (6'):

- (6') \*(C)á(C)u > HIW /ɔ:/; LTG /ɔ:/; LHI /v:/; LYP /I:/; VLW /ε:/; MTP /ε:/; LMG /ε:/; VRA /ε:/; VRS /œ:/; MSN /ɔ:/; MTA /a:/; NUM /a:/; DRG /a:/; KRO /ɛä:/; OLR /a:/; LKN /a:/; MRL /ɔ̊:/.

The same demonstration could be made for other combinations of the five protovowels – e.g. \*(C)á(C)e, \*(C)ó(C)a, \*(C)ú(C)i ...: each former trochaic foot is reflected by a single set of regular vowels similar to (4') and (6'). François (2005) lists all regular correspondences, and shows that this process of umlaut or metaphony is the key to understanding the complexification of vowel systems in the area. Overall, the historical hypothesis of metaphony provides a powerful way to disentangle the phonetic history of the area, and retrieve regularity behind modern diversity.

### 3.2 Reconstructing protoforms

By combining the regularities observed on consonants and on vowels, it is possible to reconstruct entire words. For example, if the cognate set  $\{\beta:\dots\}$  reflects a protophoneme \*β and  $\{^g\!L:r:j\dots\}$  reflects \*r, and if the hypothesis on vowels (4') is correct, then the whole cognate set for ‘stingray’ (2) can be understood as the modern result of a set of regular sound changes that affected a single protoform \*βari. Using the same procedure, one can securely reconstruct \*βati for ‘four’, and \*βaru for ‘hibiscus’.

Once a protoform has been reconstructed, it can be represented as the source of a given cognate set. Thus, whereas (2) above was just a list of modern forms, (2') entails a historical claim that all the forms are cognate, and that they reflect the specific protoform that is given as the header:

- (2') \*βari ‘stingray’ → HIW βɔ<sup>g</sup>L; LTG βer; LHI n-βæj; LYP n-βij; VLW n-βij;  
 MTP ne-βej; LMG n-βer; VRA fer; VRS βær; MSN βer; MTA βar; NUM fer; DRG βa:r;  
 KRO βeär; OLR βaj; LKN βæ:; MRL ne-βer.

The forms for ‘hibiscus’ can similarly be subsumed under \*βaru:

- (3') \*βaru ‘hibiscus’ → HIW βɔ<sup>g</sup>L; LTG βɔr; LHI n-βɔj; LYP n-βij; VLW n-βεj;  
 MTP ne-βej; LMG n-βer; VRA fer; VRS βær; MSN βɔr; MTA βar; NUM far; DRG βa:r;  
 KRO βeär; OLR βaj; LKN βa:; MRL ne-βɔr.

Each starred form constitutes a formula under which all modern forms, in spite of their current phonetic diversity, can thus be subsumed.

My current database includes about 650 lexical reconstructions; they are defined at a local level, that of “Proto Torres–Banks”. These reconstructions capture the shared history of cognate words across the various languages of the area. In some cases, the etyma I propose have also been reconstructed at a higher level in the Oceanic tree (cf. Figure 3), with earlier phonological shapes (e.g. POc \*paRi ‘stingray’, \*paRu ‘hibiscus’, \*pat ‘four’). Thus \*βari in (2') is nothing else than the form taken, in Proto Torres–Banks, by the POc etymon \*paRi. In other cases, the proposed reconstruction has no other witness outside the Torres–Banks area, and therefore represents a local innovation. In the remainder of this chapter, I will usually mention when an etymon can be traced back to earlier times; otherwise, my reconstructions will be internal to the northern Vanuatu area.

### 3.3 Avoiding false etymologies

Establishing the regular sound correspondences for each language in the area is crucial for the purpose of etymological research. This is a safe step in order to assess the cognacy between words that are potentially related, and avoid unwarranted hypotheses based on mere similarities or folk etymologies.

Let me take just one example. In Vurës, the form *wvl* has two meanings: it is a noun meaning ‘moon, month’, and it is also a geocentric directional for short distances, along vectors parallel to the shore, pointing towards Southeast. It would have been tempting to suggest a connection between the two meanings: the space directional could originate, say, in the position of the moon at certain times of the year. However, such a hypothesis is ruled out by cross-linguistic comparison, provided one pays attention to regular sound correspondences. Vurës *wvl* ‘moon, month’ belongs to the cognate set in (7). Based on regular correspondences, the best reconstruction here is a proto-form \*βula – itself a regular reflex of POc \*pulan ‘moon, month’ (Ross, Pawley, and Osmond 2003: 158):

- (7) \*βula ‘moon, month’ → VLW n-wvl; MTP nu-wvl; LMG n-wvl; VRA fulv; VRS wvl;  
 MSN wvl; MTA bla; KRO βvl; OLR βvl; LKN βvl; MRL nu-βvl.

As for the Vurës directional *wvl*, its equivalents in other languages of Vanuata Lava present different vowels. Regular vowel correspondences point to an etymon \*βolo:

- (8a) \*βolo → ‘geocentric directional for short distances parallel to the shore, pointing Southeast’: LMG wœl; VRA wvl; VRS wvl; MSN wcl.

Other northern Vanuatu languages reflect this root \*βolo as a reduplicated form \*βoloβolo, with various related meanings:

- (8b) \*βoloβolo → ‘sideways, across’: HIW wojwoj; MTA wolowolo.

- (8c) \*βoloβolo → [n] ‘a crossbeam’: MTP na-sal-wclwcl; VRS wwlwvl; LKN wclwcl.

- (8d) \*βoloβolo → [n] ‘a cross, incl. Christian cross; (astr.) the Southern Cross’: MTP nc-wclwcl; MTA wolowolo; MRL nc-wclwcl.

The best reconstruction for the latter set (8a–d) is a root \*βolo ‘crossways, across’ – and this is demonstrably the source of the directional in Vanuata Lava languages (François 2010; in prep.). It so happens that the two sequences of

syllables \*(C)ú(C)a and \*(C)ó(C)o are both regularly reflected by the same vowel /v/ in modern Vurës, which explains the homophony between *wwl* ‘moon’ (<\*βula) and *wwl* ‘Southeast’ (<\*βolo). Yet this is merely an accident of historical phonology, and any proposed connection between the two words would be a false etymology.

In sum, regular correspondences constitute a powerful touchstone when it comes to assessing the solidity of etymological hypotheses. The cautious comparison of modern forms, and the reconstruction of earlier protoforms, is the preliminary step before any reasoning about potential semantic connections becomes legitimate.

### 3.4 Reconstructing semantic change

The history of individual words sometimes involves sound change, yet little shift in meaning. For example, (7) above, reproduced below, does not involve any semantic change. Because the etymon’s reconstructable meaning is identical with that of its modern reflexes, it is indicated only once, before the arrow →:

- (7) \*βula ‘moon, month’ → VLW *n-wwl*; MTP *nu-wwl*; LMG *n-wwl*; VRA *fulu*; VRS *wwl*; MSN *wwl*; MTA *βla*; KRO *βvl*; OLR *βvl*; LKN *βvl*; MRL *nu-βvl*.

By contrast, when a shift in meaning must be reconstructed, the modern attested meaning is indicated after the arrow:

- (8a) \*βolo ‘crossways, across’ → ‘geocentric directional for short distances parallel to the shore, pointing Southeast’: LMG *wœl*; VRA *wwl*; VRS *wwl*; MSN *wɔl*.

These data sets only cite languages whose modern words reflect the protoform *and* display the meaning in question. When I only want to indicate the meaning of modern reflexes without making any claim about that of the protoform, then the gloss follows the arrow, as in (8d) below. When the semantic change also resulted in a change of syntactic category, I indicate the word’s new part of speech in brackets (e.g. [N] for noun, [v] for verb, etc.).

- (8d) \*βoloβolo → [N] ‘a cross, incl. Christian cross; (astr.) the Southern Cross’: MTP *nu-wwlwɔl*; MTA *wolowolo*; MRL *nu-wwlwɔl*.

The remainder of this study will be focused on etymologies that involve some amount of semantic change. Along with their interest for understanding

the languages and cultures of Island Melanesia, these results can feed into studies on the typology of semantic change, which are currently expanding (Traugott and Dasher 2002; Koptjevska-Tamm, Vanhove and Koch 2007; Vanhove 2008).

One methodological principle, in line with Blust (1987) or Ross, Pawley and Osmond (1998), consists in exploring a coherent semantic field so as to identify structural oppositions between different words within this domain. This will enable us, for example, to propose fine-grained nuances between words of similar meanings, such as the various etyma that can be reconstructed for designating the Underworld (§4.5.3).

As a methodological caution, I will endeavour to base every semantic reconstruction upon observable evidence. This will typically take the form of cross-linguistic comparison, whether within or without the northern Vanuatu area. The hypothesis of a semantic shift in one language can be strengthened by the existence of a similar shift, or a synchronic polysemy, in another language.

As a final illustration of the method, consider the set of forms in (9), used in modern languages to designate photographs:

- (9) ‘effigy, image, picture; (mod.) photograph’  
 MTP *nc-tɔtyal*; LMG *?ɔyal*; VRA *?ɔyal*; VRS *tɔtyial*; MTA *totoyale*; LKN *tɔtyæl*; MRL *ne-tetyel*.

The first step consists in ensuring, based on regular correspondences, that the forms are cognate: this is the case of all words in (9). Based on what we know of sound change in these languages, it is possible to reconstruct a protoform at the source of this set: this yields an etymon \*totoyale. While this etymon is unknown elsewhere,<sup>4</sup> its shape is reminiscent of two other words that can be reconstructed in northern Vanuatu: \*toto ‘chop, cut’ (POc \*tokot) and \*yale ‘tell lies, pretend’. It appears difficult to find any convincing semantic link between ‘chop’, ‘lie’ and ‘photograph’, and any semantic hypothesis in this direction would seem far-fetched – unless it was backed up by convincing evidence.

In fact, the key to this etymology is provided by the technical terminology used by carpenters. In Mwotlap, while *tɔt* means ‘chop, cut through (a piece of wood)’, the serial verb *tɔt yal*, literally ‘chop pretend’, refers to the act of notching a piece of wood – i.e. ‘cutting’ it without chopping it all the way through. Modern Mwotlap speakers explicitly point out the semantic link between this

<sup>4</sup> The words *totoyale* in Lo-Toga (Torres Is) and *totokale* in Teanu (Vanikoro, Solomon Islands; pers. data) both with the same meaning ‘image, photograph’, were evidently borrowed from Mota, at the time – late 19th century – when Mota was used by Anglican missionaries as the religious *lingua franca* of the area.

technical term for notching wood and the traditional practice of carving out an effigy in wood or treefern, to make a sculpture or human representation. Obviously, the existence of this polysemy in at least one of the modern languages is a strong argument in favour of this etymological hypothesis regarding the whole cognate set:

- (9') \**toto-yale* (lit. ‘chop falsely’) ‘notch (wood)’ → ‘carve an effigy’  
 → [N] ‘effigy, image, picture; (mod.) photograph’  
 MTP *nɔ-tɔtyal*; LMG *?ɔyal*; VRA *?vyal*; VRS *tɔtyal*; MTA *totoyale*; LKN *tɔtyæl*; MRL *ne-tetyel*.

The phonological form of the reconstruction is given for the level of “Proto Torres–Banks”, the putative protolanguage from which modern reflexes can be derived. However, I make no claim about the historical depth of such a reconstruction. It is quite possible that the compound \**totoyale* is a relatively recent coinage, dating back perhaps just a few centuries, which spread through a portion of the dialect continuum. I also make no claim about the date of the semantic change (from a verb ‘notch’ to a noun ‘image’), which could well have taken place recently and diffused across languages via contact. What I am really interested in is the fact that a set of modern forms can be safely shown to descend from a single etymon; and that the meanings associated with that cognate set can be shown to have evolved in a certain way during the history of these languages. In other terms, the main objective of this study is not so much to reconstruct a particular protolanguage (“Proto Torres–Banks”), as it is to unravel the history of individual words.

## 4 The history of spiritual words in northern Vanuatu

The previous section outlined the methodological principles that underlie the quest for reliable etymologies. The two pillars of this research are the solidity of sound correspondences, and the empirically assessed plausibility of semantic changes.

Based on these methodological principles, I here propose to explore the spiritual world of northern Vanuatu, via the words – and their derivatives – that relate to notions of ‘person’, ‘mind’, ‘soul’, ‘ghost’ and other kinds of ‘spirits’. Overall, I will mention about 80 concepts, providing each with a protoform and, whenever possible, with etymological comments.

For reasons of space, I won’t be able to mention all relevant words of modern languages, and will reduce my selection to those reconstructable words that are most significant, most revealing, or best represented in the Torres–Banks languages. Likewise, I will not be able to delve into all the ethnographic detail which would be ideally involved in dealing with such a rich domain. I shall restrict myself to those issues for which the linguist can make a contribution, and refer the interested reader to the abundant anthropological literature that revolves around the spiritual world of northern Vanuatu cultures – see, *inter alia*, Codrington (1891), Ivens (1931), Vienne (1984), Jolly (1996), Lanouguère-Bruneau (2002), Mondragón (2004), Hess (2009).

### 4.1 The person

#### 4.1.1 Terms for ‘person’

Along with the gender-specific terms for ‘man’ (see [16]–[17]) and ‘woman’ (see [1] above), Northern Vanuatu languages have a generic term for ‘person (of unspecified gender)’:

- (10) ‘individual, person, human being’:  
 HIW *tajo*; LTG *telə*; LHI *n-at*; LYP *n-at*; VLW *n-at*; MTP *n-et*; LMG *n-?vɔŋsvr*; VRA *?taŋsara*; VRS *taŋsar*; MSN *taŋmʷsar*; MTA *tanun*; NUM *tu"duₙ*; DRG *t"duₙ*; KRO *tu"duₙ*; OLR *tufun ~ tʃufun ~ tuʃuŋ ~ tʃuʃuŋ*; LKN *fə:fun*; MRL *ne-te"duₙ*.

I will propose that all the forms in (10) ultimately derive from the same root, namely \**ata* ‘person’. Four languages preserve this etymon as such, in their word for ‘person’:

- (11) \**ata* ‘individual, person, human being’ → LHI *n-at*; LYP *n-at*; VLW *n-at*; MTP *n-et*

This root \**ata* is the same word as POc \**qata* (Ross, Pawley and Osmond, in prep. b) and PNCV \**?ata* (Clark 2009: 76), which have been reconstructed with the same meaning ‘person’. It is different from the homophonous root \**qata* meaning ‘soul, spirit (of s.o.)’ which I will discuss in §4.3.

Torres–Banks languages have a number of nouns with human referents whose first element is \**ta-*. In principle, this \**ta-* prefix could reflect POc \**tau*, which also meant ‘person’. But this is challenged by the observation that some derived forms reflect the retention of a vowel /a/ before the /t/ – e.g. Lehali

*atŋʷan* or Vurës *atŋʷmʷin* ‘man’ as in example (16) below, or Lakon *ætmæt* ‘ghost’ (44), etc. This suggests that the apparent \**ta-* prefix should really be reconstructed as \**ata-*, with reflexes showing frequent apheresis and sporadic retention. Because the retention of the first vowel in \**ata* is the exception rather than the rule, I will henceforth represent the prefix as \*[*a*]*ta*.<sup>5</sup> However, etymologically speaking, it is clear that this \*[*a*]*ta*- prefix is the same root as the noun (11) \**ata* ‘person’.

This prefical element \*[*a*]*ta*- is found in the other forms of (10) with the generic meaning ‘person’. Thus, several languages reflect an etymon which reconstructs as \**ta*<sup>n</sup>*dunu*:

- (12) \**ta*<sup>n</sup>*dunu* → ‘individual, person, human being’:

VRS *tø*<sup>n</sup>*dün*; MTA *tanun*; NUM *tu*<sup>n</sup>*dun*; DRG *t*<sup>n</sup>*dun*; KRO *tu*<sup>n</sup>*dun*; OLR *tufun* ~ *tfufun* ~ *tufuŋ* ~ *tfufuŋ*; LKN *fa:fun*; MRL *ne-te*<sup>n</sup>*dun*.

Even though the word has become unanalysable in modern languages, it was originally a compound form \*[*a*]*ta*-*dunu*, in which the second element \**n**dunu* meant ‘true’:

- (13) \**n**dunu* ‘true, real’ → VRS *n**dün*; MSN *nun*; MTA *nun*; NUM *βi*<sup>n</sup>*dun*; DRG *le*<sup>n</sup>*dun*

As pointed out by Codrington and Palmer (1896: 195), this etymon \*[*a*]*ta*-*n**dunu* ‘real person’ stands in contrast with \*[*a*]*ta-mate* ‘dead person, ghost’ (§4.4.1). In this part of the world, what defines a ‘human being’ is not the contrast with animals, but first and foremost its difference from ghosts and evil spirits. Note this is distinct from the common use, in other parts of the world, of ‘real, true’ as an ethnonym (e.g. *Māori* < *māori* ‘true’): the latter pattern draws a contrast between the ‘true’ members of a social group and other human groups – a notion that is clearly absent from the meaning of \*[*a*]*ta*-*n**dunu*.

The four languages of Vanua Lava reflect a different protoform for ‘person’:

- (14) \**ta*<sup>w</sup>*asara* → ‘individual, person, human being’:

LMG *n-?vŋsvr*; VRA *?aŋsara*; VRS *tansar*; MSN *taŋmʷsar*

<sup>5</sup> The use of square brackets follows established usage among Oceanicists (cf. Ross, Pawley and Osmond 1998: 13). Whereas round brackets refer to protophonemes whose presence in the etymon cannot be determined, square brackets symbolize the fact that two distinct reconstructions, one with and one without the bracketed element, can both be independently reconstructed based on modern evidence, and are equally robust.

This form \**ta*-*m*<sup>w</sup>*asara* is also a compound, with the same base \*[*a*]*ta* ‘person’. The second element \**m*<sup>w</sup>*asara* is reminiscent of an adjective whose modern meaning is ‘poor, needy’:

- (15) \**m*<sup>w</sup>*asara* ‘poor, needy’ → LHI *ŋʷəsaŋ*; LYP *ŋmʷasaj*; LMG *ŋmʷvŋsvr*; MTP *ŋmʷasaj*; MTA *ŋmʷasara*; DRG *ŋmʷsar*; LKN *ŋmʷasa*:

Why would a compound ‘poor person’ become the normal word for ‘person’? A tentative hypothesis could be that \**m*<sup>w</sup>*asara* had a social meaning like ‘commoner’, and contrasted with a higher social class. Indeed, we will see that the antonym of \**m*<sup>w</sup>*asara*, namely \**m*<sup>w</sup>*era-i somu* ‘rich’, has lexical ties with the former political system, and with a class of highly ranked chiefs who were regarded as supernatural (§4.7.1). By contrast, \**ta*-*m*<sup>w</sup>*asara* (lit. ‘poor person’) would have meant ‘commoner, mundane person’, before generalising as the default term for ‘person’.

The two Torres languages show two forms (HIW *taje*; LTG *tela*) which seem to reflect a protoform \**tala* or \**talua*. While these seem to include the same \**ta*-prefix, the etymology of the whole compound is unclear (but see fn. 9).

#### 4.1.2 The question of gender

The element \*[*a*]*ta*, which we just saw in compounds with the generic meaning ‘person’, can also be detected in various compounds referring to humans. Some of these compounds raise the question of whether \**ata* was really unspecified for gender, or typically associated with male referents.

The word for ‘male, man, husband’ can locally be reconstructed as a proto-form \*[*a*]*tam*<sup>w</sup>*ane*:

- (16) \*[*a*]*tam*<sup>w</sup>*ane* → ‘male, man, husband’:

HIW *tøŋʷen*; LTG *tøŋʷen*; LHI *n-atŋʷan*; LYP *n-taŋmʷan*; VLW *n-taŋmʷan*; MTP *na-tjŋʷan*; LMG *n-?aiŋmʷan*; VRA *?aŋmʷan*; VRS *atŋmʷm*; MSN *atŋmʷm*.

This form reflects a PNCV compound \*(?)*ata-m*<sup>w</sup>*a?**ane* (Clark 2009: 76), in which the second element reflects POc \**maRuqane* ‘male’ (Blust 2012) or its doublet \**m*<sup>w</sup>*aqane* (Ross, Pawley and Osmond in prep.). In this word, the gender information can be attributed to this second element, and the prefix \*[*a*]*ta*- may still be analysed as gender-neutral: \*[*a*]*ta*-*m*<sup>w</sup>*ane* literally parses as ‘person-male’.

The situation is a bit different in the southern Banks languages, which have lost \*[*a*]*ta*-*m*<sup>w</sup>*ane*. They have replaced it with a new compound (17)

\**mʷera-i ata*, in which the second element is again \**ata* ‘person’. As for the first element, it is a noun \**mʷera* ‘child’ (François 2011a: 185; Clark 2009: 155). Literally, the compound \**mʷereata* can be interpreted as ‘child of a person’; it likely referred originally to a ‘boy child’ (opp. ‘girl’), and later generalized as ‘male, man’.

- (17) \**mʷera-i ata* ‘child of a person’ → \**mʷereata* ‘male, man, husband’: MTA *ŋmʷereata*; NUM *ŋmʷerat*; DRG *ŋmʷerat*; KRO *ŋmʷarat*; OLR *ŋmʷarat*; LKN *ŋmʷεrat*; MRL *na-ŋwarat*.

In Mota, the singular form *ŋmʷereata* ‘man’ has a suppletive form for plural: *rereata* ‘men’ – itself a compound of *rere* ‘Plural’ and \**ata* ‘person’. It is worth of notice, again, that phrases that are apparently unspecified for gender, such as ‘child of a person’, or ‘PLURAL:person’, have lexicalized with specifically male referents. In the case of \**mʷereata*, the gender shift is to be assigned not so much to \**ata* as to the first element \**mʷera* ‘child’ → ‘boy child, son’.

It is not the case, however, that all reflexes of \**ata* are specifically male. The prefix \*[*a*]ta- is also found on the Mota form for ‘woman, female’, namely *taβne* – from \**ta-βine* ‘person–woman’ [see (1) above]. Lo-Toga has a reflex of the same form but with a different meaning, *taβinə* ‘mother-in-law’. Both reflect a POc compound \**qata-pine* ‘woman, female’, from \*(*pa*)*pine* with the same meaning (Ross, Pawley and Osmond, in prep. b).

The gender ambiguity is also observable in the pronoun forms that were created in some languages, and which also contain \**ata*. The two Torres languages have thus developed a set of pronouns that code for number and gender. Even though \**ata*, as an etymon, is normally gender-neutral, it is only present in the masculine forms of these pronouns: HIW *te⁹L\atə* = LTG *hori\atə* (masculine dual) ‘two men’.

Conversely, Lakon has a small paradigm of anaphoric pronouns that include \**ata* as a second element (the first element is the 3rd person pronoun): *worw-at* ‘these two people’, *tlti:-at* ‘these three people’, *yr:-at* ‘these people’. Although these forms are in principle neutral in gender, they contrast with another set of pronouns (*worw-tr:*, *tlti:-tr:*, *yr:-tr:*) that are specifically masculine. As a consequence of this asymmetry, the neutral forms ending in -*at* tend to be more frequently associated with female referents.

In sum, the root \**ata* must be reconstructed as neutral in terms of gender, even though some its reflexes have become specifically associated with male or with female referents.

#### 4.1.3 Other terms derived from \**ata* ‘person’

Northern Vanuatu languages have a noun \**tamarayai* referring to an ‘old man’. As a noun, this word is normally specifically male – contrasting with \**mayatea* ‘old woman’ – but as an adjective, it can be used for any gender, meaning ‘old’:

- (18) \**tamarayai* → ‘old man; old’: HIW *tama⁹Le⁹Le*; LHI *tamajyæ*; LYP *tamajye*; VLW *tamajye*; MTP *tamajye*; VRA *ʔamaryā*; VRS *tamarye*; MSN *tamarye*; MTA *tamarayai*; NUM *tamaraya*; DRG *tmarya*; LKN *tama:yæ*; MRL *temerye*

This word is an interesting compound, based on the \*[*a*]ta- element. The second element is reminiscent of a word that has a very different meaning, namely ‘shy, ashamed’. The semantic connection is opaque, until one realizes that the words in (19) did not initially mean ‘shy’, but were really a verb ‘tremble, quiver’ (Codrington and Palmer 1896: 72):

- (19) \*[*ma*]*marayai* ‘tremble, quiver’ → ‘shy, ashamed, shameful’: LHI *mmajyæ*; LYP *mmajyε*; MTP *mamajyε*; VRA *marya*; VRS *ia\marye*; MTA *marayai*; NUM *maraya*; DRG *marya*; LKN *ma:yæ*

An old man (18) is thus \*[*a*]ta-*marayai* ‘a quivering person’. As for the notion of shyness, it is in fact a construction in which the subject is a noun \**a⁹be* ‘body’; so, ‘I feel shy’ is literally ‘my body is quivering’. Finally, the same root \**marayai* ‘quiver’ can be found in Mwotlap *mat-majye* ‘sleepiness’ – etymologically \**mata-marayai* ‘quivering eyes’.

Another word referring to an ‘old person’ (of any gender) is \**tamatua*:

- (20) \**tamatua* → ‘old person’: HIW *taməso*; LTG *teməto*; MTA *tamatua*

This word is a compound of \*[*a*]ta- ‘person’ and \**matua* ‘ripe, mature’ (> HIW *məso* ‘ripe, fullgrown; large’) – itself from POc \**matuqa* ‘ripe, mature, fullgrown’. The metaphorical connection between a ripe fruit and old age is evidently an ancient one, as it is reflected in many Oceanic languages (Ross, Pawley and Osmond 2003: 204).

I will later cite the compound (86) \**ta-βusi-mʷele* ‘man of high rank’ (§4.7.2), also derived from \*[*a*]ta-. Finally, a last compound based on that root is \*[*a*]ta-*mate* ‘dead person, ghost, ancestral spirit’. This important word will be the focus of §4.4 below.

## 4.2 The mind

The notion of someone's mind or psyche is covered by a number of different terms in northern Vanuatu languages, and I will only cite a few of them here.

Sometimes, the word *\*mb<sup>w</sup>atu* 'head' (< POC *\*mb<sup>w</sup>atu*) is used to refer to one's intelligence. I here give the form with the 1sg possessor (\*-<sup>w</sup>gu):

- (21) *\*mb<sup>w</sup>atu* (-<sup>w</sup>gu) '(my) head' → '(my) head; (my) intelligence':  
 HIW *k<sup>w</sup>iti-k*; LTG *k<sup>w</sup>ətu-k*; LHI *n-k<sup>w</sup>utu-k*; LYP *n-k<sup>w</sup>ifili-k*; VLW *n-ŋg<sup>w</sup>iti-ŋ*  
 MTP *ni-k<sup>w</sup>ti-k*; LMG *n-k<sup>w</sup>iʔi-k*; VRA *k<sup>w</sup>iʔi-k*; VRS *k<sup>w</sup>ətū-k*; MSN *k<sup>w</sup>utu-k*  
 MTA *k<sup>w</sup>atu-k*; NUM *na-k<sup>w</sup>utu-k*; DRG *na-k<sup>w</sup>tu-k*; KRO *k<sup>w</sup>utu-k*; OLR *k<sup>w</sup>utu-k*  
 LKN *k<sup>w</sup>atv-k*; MRL *na-k<sup>w</sup>atv-k*

An example would be Mwotlap *ni-k<sup>w</sup>ti ne-hej* 'you have a sharp mind' (lit. 'your head is sharp'); or *ni-k<sup>w</sup>ti na-wak* 'you're open-minded, you have a curious mind' (lit. 'your head is open'). For these intellectual senses, Mwotlap can also use *nv-svl* 'brain' and *ni-mma-n* 'fontanelle', yet these do not seem to be frequently used in other languages of the region. Torres–Banks languages occasionally describe certain feelings using other body parts – such as the 'diaphragm' (*\*mwala*) for surprise; the 'belly' (*\*to<sup>m</sup>b<sup>w</sup>a*) for desire; the 'liver' (*\*βara*) for awe and fear. For reasons of length, I shall not illustrate these senses here.

The most common term when referring to the mind and to certain emotions are reflexes of an inalienable noun which can be reconstructed as *\*lolo*. With a 1sg possessor, the suffixed form reconstructs as *\*lolo-<sup>w</sup>gu*:<sup>6</sup>

- (22) *\*lolo* (-<sup>w</sup>gu) → '(my) mind, will...':  
 HIW *je-k*; LTG *lio-k*; LHI *n-lr-k*; LYP *n-lø-k*; VLW *n-lr-ŋ*; MTP *na-li-k*; LMG *n-lø-k*  
 VRA *lv-k*; VRS *lølø-k*; MSN *lvlv-k*; MTA *lolo-k*; NUM *na-llo-k*; DRG *na-llo-k*  
 KRO *na-llo-k*; OLR *lvlv-k*; LKN *lølø-k*; MRL *lo-k*

This noun is slightly atypical. While it behaves morphologically like a body part (obligatory presence of possessor), it cannot be located anywhere on the body. It never appears in the position of object or adjunct (as in Eng. *in my mind*), and is restricted to subject position. Furthermore, in each language it can only take a small number of predicates. For example, in Mwotlap, *na-li-k* is only used as

the subject of *m[ε]jen* 'daylight' and *k<sup>w</sup>vŋ* 'night', in constructions that mean respectively 'remember; be lucid' and 'forget; be unconscious':<sup>7</sup>

- (23) MTP *na-le-k*      *me-myen*      *ēgēn*.  
 /na-li-k      me-mjen      ryin/  
 ART-mind-1sg PRF-daylight now  
 'I remember now.'

- (24) MTP *na-lo-n*      *may qōñ*.  
 /na-lo-n      maj k<sup>w</sup>vŋ/  
 ART-mind-3sg CPLT night  
 'He has already forgotten (it)' ~ 'He's unconscious' ~ 'He's senile.'

Given these sentences, a language-internal analysis of Mwotlap may define *na-li-k* as 'memory' or 'consciousness'. And indeed, modern speakers have no other gloss for it: the *na-li-k* is simply 'that thing in you that gets dark when you forget, and illuminated when you remember'. Notice, in passing, the metaphor of light that is also found in the etymologies of English *lucid*, *enlighten*, *illuminate* or *brilliant*.

But the observation of cognate forms in languages other than Mwotlap makes it clear that the underlying meaning of *\*lolo* must be broader than just 'memory'. While they all share with Mwotlap the constructions illustrated in (23)–(24), they also use the forms in (22) in constructions referring to anger. Depending on the language, the possessed noun is then the subject of an adjectival predicate *\*saat[i]* 'bad' (Lemerig in [25]); or of a verb *\*anu* 'sting, irritate' (Vurës in [26]):

- (25) LMG *lö-k*      *se’ sā n-meri now*.  
 /lø-k      se? sv n-ŋm<sup>w</sup>eri nɔw/  
 mind-1sg bad FOC<sub>1</sub> ART-child that:FOC<sub>2</sub>  
 'I'm angry at that child' (lit. 'My mind is bad (at) that child')

<sup>6</sup> The simplification of geminates (e.g. *\*nà loló-<sup>w</sup>gu* > *\*na-llo-k* > *na-li-k*) is regular in Mwotlap (François 2000a: 59–60) and other northern languages.

<sup>7</sup> Sentential examples in individual languages are provided in standard orthographies, followed by a phonemic transcription. Abbreviations in glosses include: ART article; CPLZR complementizer; CPLT complete aspect; DEP dependent clause; FOC focus marker; FUT future; POT potential; PRF perfect; PROH prohibitive.

- (26) VRS *nitog na lolo-ñ ge en no!*  
 /nitɔg na lɔlɔ-ñ ðe en nɔ/  
 PROH ART mind-2sg FUT sting 1sg  
 ‘Don’t be mad at me!’  
 [lit. ‘Don’t let your mind sting me’]

Somewhat more figuratively, several languages use a verb \**ngoto* ‘nip, bite’ (Lakon in [27]) or \**yara* ‘bite’ (Mwesen in [28]):

- (27) LKN *lolo-ñ mētē kot lee na!*  
 /lɔlɔ-ñ miti kɔt le: na/  
 mind-2sg PROH<sub>1</sub> nip PROH<sub>2</sub> 1sg  
 ‘Don’t be mad at me!’  
 [lit. ‘Don’t let your mind nip me!']
- (28) MSN *lōlō-k mete gagar le nēk.*  
 /lʊlʊ-k mete yayar le nɪk/  
 mind-1sg NEG:POT<sub>1</sub> bite POT<sub>2</sub> 2sg  
 ‘I won’t be angry at you.’  
 [lit. ‘My mind won’t bite you’]

Finally, the two Torres languages employ \**lolo* as the subject of a verb ‘want, like’ – as in Hiw (29):

- (29) Hiw *ne yō-k ñis tom ke tō veñog.*  
 /nə jø-k ñis tɔm kə to βəg̊lɔy/  
 ART mind-1sg want CPLZR 1sg:DEP go:SG also  
 ‘I want to go too.’ [lit. ‘My mind wants that I go.’]

In sum, the meaning of \**lolo* includes not only reference to the ‘memory’, but also to the seat of will and emotions – at least anger. The abstract gloss ‘mind’ reflects this semantic versatility.

Crucially, it is possible to identify the etymology of the word \**lolo* which is here glossed ‘mind’. This is historically the same word as \**lolo* ‘inner part, interior’ which otherwise designates the inside of an inanimate object, e.g. a house. Because this spatial use of ‘inside’ is normally incompatible with human possessors, speakers of at least some languages perceive the two constructions as involving separate words. For instance, my Mwotlap consultants never drew a connection between /lɔ/ in *na-lɔ iñm̊w* ‘the interior of the house’ and /lɪ/ in (23)

*na-li-k* ‘my memory’. This is partly due to morphological processes of ablaut – e.g. 1sg /lɪ-k/, 3sg /lɔ-n/ – which have blurred the link between otherwise cognate forms (François 2005: 485). Only cross-linguistic comparison made it obvious that the word for ‘mind’ is really the same word as ‘interior’ – via a metaphor we have in English with *inner thoughts* or *inner man*. The same polysemy is found elsewhere in the Austronesian family, whether associated with the same POc etymon \**lalo[m]* ‘inside, inner part; mind, feelings’ like here in northern Vanuatu; or with POc \**loto* ‘inner self, feelings, mind’ (Blust 2012).

Finally, the identification of an etymon \**lolo* ‘interior, inside (of s.th.); mind (of s.o.)’ makes it possible to unravel the etymology of a whole set of compounds based on this noun. They all follow a pattern in which \**lolo* is an incorporated subject, followed by its predicate; the output of this composition can be a verb, an adjective or a noun. For example, \**lolo-* ‘interior, mind’ and \**mbwɔŋi* ‘night’ (<POc \**mboŋi*) form together a compound \**lolo-mbwɔŋi* (lit. ‘mind in night’). As a verb, the form means ‘forget’. As an adjective, it translates as ‘stupid, ignorant’ (cf. Eng. *benighted*) or sometimes ‘crazy’. As a noun, it refers to ‘heathen times’ before Christianisation.

- (30) \**lolo-mbwɔŋi* ‘mind in night’  
 → ‘[v] forget; [adj] stupid, ignorant, foolish; [n] heathen times’:  
 HIW *jejkəŋ*; VLW *lɔl̩ŋgbẘŋ*; MTP *lɔlk̪ẘŋ*; LMG *lilk̪ẘŋ*; VRS (*lwl̩lk̪ẘŋ*);  
 MSN (*lɔlɔsil*); MTA *lolok̪ẘŋ*; LKN (*lɔlhil*); MRL (*lɔlsil*)<sup>8</sup>

Thus (24') is essentially an equivalent of (24) above, in the same language Mwotlap:

- (24') MTP *kē may lol-qōñ.*  
 /kɪ maj lɔl-kp̊ẘŋ/  
 3sg CPLT mind-night  
 ‘He has already forgotten (it).’

The antonym of \**lolo-mbwɔŋi* is \**lolo-* + \**marani* ‘daylight’ (<POc \**raqan* ‘day’). The sentence (23) above already illustrated how the noun \**lolo-* can be the subject of a predicate \**marani* ‘daylight’: the phrase resulted in a meaning ‘remember’. The compound \**lolo-marani*, although based on the same elements, takes up a meaning of its own:

<sup>8</sup> Forms in brackets are compounds that follow the same pattern, but have replaced the second element (\**mbwɔŋi* ‘night’) with a synonym: -*sil* or -*hil* < \**sili* ‘dark’; -*kpẘor* < \**mbwure* ‘ignorant’.

- (31a) \**lolo-marani* ‘mind in daylight’  
 → ‘[ADJ] wise, intelligent, knowledgeable; [N] wisdom, knowledge’: LHI *ləlməjæn*; LYP *ləlməjen*; VLA *ləlməjen*; MTP *ləlməjen*; LMG *ləlməren*; VRA *ləlməran*; VRS *ləlməren*; MSN *ləlməren*; MTA *lolomaran*; DRG *ləlməra:n*; LKN *ləlmə:rən*; MRL *ləlməren*

In the two Torres languages, the reflexes of this etymon have become the verb ‘know’:

- (31b) \**lolo-marani* ‘mind in daylight’  
 → ‘[ADJ] wise, intelligent, knowledgeable; [N] wisdom, knowledge; [V] know’: HIW *jəjmə:glen*; LTG *lolməren*

The semantic differences between the phrasal constructions and the lexical compounds are summarized in *Table 1*.

	* <i>lolo</i> ‘mind’ + * <i>bwɔŋi</i> ‘night’	* <i>lolo</i> ‘mind’ + * <i>marani</i> ‘daylight’
PHRASE (Subject + Predicate)	forget; be unconscious; be senile	remember; be lucid
COMPOUND	[V] forget; [ADJ] ignorant, foolish; [N] heathen times	[ADJ] wise, intelligent, knowledgeable; [N] wisdom, knowledge; [V] know

Table 1: Meaning of some phrases vs compounds involving \**lolo* ‘inside; mind’

One also finds various \**lolo-* compounds referring to anger:

- (32) \**lolo-saat[i]* ‘mind bad’ → ‘[ADJ] angry; [N] anger’: LTG *ləlhia*; LHI *ləlsæt*; LYP *ləlsit*; LMG *ləlse?*; VRA *ləlse?*  
 (33) \**lolo-ŋoto* ‘mind nip’ → ‘[ADJ] angry; [N] anger’: DRG *ləlkətkət*; ORL *ləlkətkət*; LKN *ləlkət*  
 (34) \**lolo-yara* ‘mind bite’ → ‘[ADJ] irritated, angry; [N] anger, irritation’: MTP *ləlyayaj*; VRS *ləlyayar*; MTA *loloyara*  
 (35) \**lolo-anu* ‘mind sting’ → ‘[ADJ] irritated, angry; [N] anger, irritation’: VRS *ləlenen*; MTA *loloanu*; MRL *lələn*

Finally, the languages of the area have two other compounds:

- (36) \**lolo-βono* ‘mind blocked’  
 → ‘[ADJ] sad, sorrowful, melancholic; [N] sorrow’: LHI *ləlwən*; VLA *ləlwən*; MTP *ləlwən*; MTP *ləlwən*; VRS *ləlwən*; MSN *ləlwən*; MTA *lolowono*; NUM *ləlwən*; DRG *ləlwən*; LKN *ləlwən*; MRL *ləlwən*  
 (37) \**lolo-wia* ‘mind good’ → ‘[ADJ] kind, thankful; [INTERJ] thanks’: HIW *jəjwja*; LTG *lowia*; MTA *lolowia*; MRL *ləlwēa*

Much more could be said on the lexical domain of psychology and feelings, but this would require a detailed semantic analysis of each language separately, and would take us beyond the scope of this study. Suffice it to say that the most widespread etymon referring to the ‘mind’ in northern Vanuatu is \**lolo* – etymologically ‘inside’.

### 4.3 The soul

The term \**lolo* ‘inside; mind’ is only used about living, sentient individuals. The same languages employ different, unrelated words to express the notion of ‘soul’ – that is, the living part of a person separable from the body. The domain of the soul will lead us to explore a new territory – namely, the world of ghosts and spirits.

Importantly, the words cited in this section are never used to refer to the mind or psyche of a living individual, as in English *deep in my soul* or *a sensitive soul*. Contrary to many languages in which these notions are etymologically related,<sup>9</sup> ‘mind’ and ‘soul’ are two separate notions in Vanuatu languages, which have nothing in common, and are never used in the same contexts. The notion of ‘soul’ is mentioned exclusively in those contexts where the living part of an individual is said to separate from their body, and move to other locations.

The cultures of northern Vanuatu mention three kinds of contexts when this sort of separation takes place. First, when somebody dies, their soul will separate from their body, and hover close to it for about five days (Hess 2009: 161), until it starts its voyage towards the Other World (see §4.5). Another occasion is when a victim of sorcery becomes deeply sick or insane: this is because their soul has been kidnapped by evil spirits, and needs to be brought back to their body. The third occasion when souls are mentioned corresponds precisely to the magic

<sup>9</sup> Cf. Latin *animus* and *anima*, or the various senses of *spīritus*, of Sanskrit *ātman*, Arabic *rūh*, etc. (François 2008b: 187–213).

practice of healers or shamans: they are the ones who can get their own soul to wander between worlds while their body is asleep. This power allows them, for example, to travel across to the Other World, and bring back the soul of the sick person they want to heal (see [40], [77] below).

#### 4.3.1 POc \*qata: one or two roots?

In several languages, the word for ‘soul’ is a noun *\*ata* – a homophone with the other word *\*ata* ‘person, human being’ we saw earlier (11). One morphological difference, though, is that *\*ata* ‘soul’ is an inalienable (relational) concept, which obligatorily indexes a possessor (‘soul of s.o.’). The latter normally takes the form of a possessive suffix, which is here given in its 3sg form *\*-na*:

- (38) *\*ata (-na)* ‘(his/her) soul, spirit’  
 → HIW *ata-na*; LTG *te-na*; LHI *n-eta-n*; VRS *ata-n*; MTA *ata-na*.

In terms of historical morphology, it is relevant to note that the suffixation entailed a shift in stress, which occasionally altered the quality of some vowels. Thus for the single language Lehalı, compare *\*(na) áta > n-at* ‘person’ with *\*(na) atá-na > n-eta-n* ‘his/her soul’. In Vurès, former processes of umlaut (see §3.1), combined with vowel harmony, have triggered stem alternations: *\*atá-ŋgu > vrs cete-k* ‘my soul’; *\*atá-na > ata-n* ‘his/her soul’; *\*atá-i > ete* ‘soul of’ (François 2005: 484–488). As for the sequence (C)e(C)a in Lo-Toga, it is the regular reflex of an earlier string *\*(C)a(C)a* (François 2005: 490).

Proto Torres–Banks *\*ata* ‘soul, spirit’ is the same word as PNCV *\*?ata*, whose sense ‘soul, spirit’ is found in other languages of Vanuatu (Clark 2009: 76). The three languages of Vanikoro, in the Solomon Is., also reflect it: Teanu *ata*, Lovono *ala*, Tanema *ae* ‘soul, spirit’ (François 2009: 111). Cognate forms in other Oceanic languages have a meaning ‘shadow, reflection’, as in Tongan *?ata* (Ross, Pawley and Osmond, in prep.) – a sense which is lexified differently in north Vanuatu. All this points to the reconstruction of a POc form *\*qata* ‘shadow, reflection; soul, spirit’.

It is not entirely clear whether POc *\*qata* ‘person, human being’ and POc *\*qata* ‘shadow, reflection; soul’ formed a case of polysemy of a single word – or if they were mere homophones. The semantic connection between ‘person’ and ‘soul, essential living component of a person’ does not seem implausible, and one could propose that the two roots may be ultimately the same word. However, a discussion with several specialists of the domain (R. Blust; A. Pawley; M. Ross; M. Osmond, p.c.) yielded the conclusion that POc *\*qata* is more

probably a case of homophonous terms arising from different sources (see also Osmond, this volume). *Table 2* sums up the steps that can be reconstructed for the shifts in meaning and sound of the two roots, which are further commented upon below.

	PAN	Pre-POc	POc	PNCV	Torres-Banks	e.g. Lehalı
‘outsider, alien’	<i>*qaRta &gt;</i>	<i>*qata</i>				
		↓				
‘person’		<i>*qata &gt;</i>	<i>*qata &gt;</i>	<i>*?ata &gt;</i>	<i>*ata &gt;</i>	(11) <i>n-at</i>
‘shadow, reflection’	<i>*qantad</i>	<i>&gt;</i>	<i>*qata</i>			
			↓			
‘soul, spirit’		<i>*qata &gt;</i>	<i>*?ata &gt;</i>	<i>*ata &gt;</i>		(38) <i>n-eta-n</i>

Table 2: The two homophones *\*qata* of Proto Oceanic

The starting point for the first etymon is a Proto Malayo Polynesian root *\*qaRta*, which Blust (1972, 2012) reconstructs as ‘outsider, alien’; it shows a shift to ‘human being, person’ even before Proto Oceanic, in various Central Malayo-Polynesian languages (Blust 2012). At the level of POc, the form had changed to *\*qata*, and the original meaning ‘outsiders, alien people’ had evidently been lost, as it is attested nowhere in Oceanic languages; the word had broadened its meaning to ‘person, human being’, for which *\*qata* then competed with *\*tau* (Ross, Pawley and Osmond, in prep.). The Banks and Torres forms cited in (11), illustrated in *Table 2* with Lehalı, reflect regular sound change from *\*qata*.

As for the second *\*qata*, it has been linked (Ross, Pawley and Osmond, in prep.) with a PAN root *\*qantad* ‘shadow, reflection, image, likeness’ (Dahl 1981). The semantic connection between ‘soul, spirit’ and ‘shadow’ is relatively widespread in the Austronesian word, whether in relation to this root *\*qata* or with POc *\*qanunu* ‘shadow, reflection, soul’ (Blust 2012). Common to all these notions is the reference to a person’s particular “presence”, when it contrasts with the physical reality of their body. The spirits of our ancestors are present amongst us, in our personal memories and in our collective representations – like the shadows of bygone lives.

#### 4.3.2 A second etymon for ‘soul’

Besides the form *\*qata* given in (38), some Banks languages lexify ‘soul, spirit’ with a different protoform *\*tala (-na)*:

- (39) \**tala* (-na) → ‘(his/her) soul, spirit’  
 LYP *n-tala-n*; VLW *n-tala-n*; MTP *na-tala-n*; LMG *n-?alb-n*; LKN *tala/tri-n*

The origin of \**tala* is unclear. It may ultimately be derived from \**ata* (?), plus an element \*-la- of unknown origin.<sup>10</sup>

My Mwotlap informants were specific about what has a soul and what does not. Plants, as well as smaller animals such as insects, mice, or fish, are not endowed with a soul. By contrast, higher animates have a soul, including pigs (an important animal in traditional society), whales, dolphins, and the dreaded sea-snakes – but not sharks or turtles.<sup>11</sup>

This root \**tala* can be found in the compound \**tala-mauri* – literally ‘soul alive’ (<POc \**maqurip* ‘live, alive’). This refers to a magical practice whereby a person endowed with supernatural powers (\**mana*, see (90) below) is capable, while their body is asleep, of letting their own soul migrate and wander between worlds, for a few hours, before returning to the body. This practice can have a specific purpose – e.g. the magician is a healer who needs to migrate to the realm of the dead, in order to fetch the soul of a sick person – or may be purely gratuitous.<sup>12</sup>

- (40a) \**tala-mauri* ‘soul alive’ → ‘a magical practice whereby the soul of a living healer wanders away from their body’: MTP *na-talmij*

Codrington and Palmer (1896: 189) has a slightly different definition for the cognate form in Mota:

- (40b) \**tala-mauri* → ‘one whose soul goes out to eat the soul of a dead person’: MTA *talamaur*

The etymology ‘soul alive’ is presumably grounded on the fact that this magical practice is carried out by a living person, who has the power to bring the soul back to the body; this is in contrast with the final migration of one’s soul out of the body, which follows death. The realm of souls and ghosts is precisely the object of the next section.

<sup>10</sup> It is possible, though hard to confirm, that this \**tala* may be the same form underlying the Lo-Toga noun for ‘person’, *telə* < \**tala* (§4.1.1).

<sup>11</sup> Note, however, that reports sometimes vary on these matters (Lanouguère-Bruneau 2002: 220).

<sup>12</sup> We will see below a related notion \*[*a*]tamate-rojo (77).

## 4.4 The names of ghosts and spirits

### 4.4.1 On life and death

The stages of life and death are expressed respectively by two verbs: \**esu* and \**mate*.

- (41) \**esu* ‘live, be alive’ → HIW (*at*); LTG (*ah*); LHI *ih*; LYP *IS*; VLW *Ih*; MTP *ih*; LMG *IS*; VRA *IS*; VRS *IS*; MSN *IS*; MTA *es*; NUM *IS*; DRG *IS*; KRO *IS*; OLR *IS*; LKN *IS*; MRL *IS*

This form \**esu* is a local innovation; the POc etymon \**maqurip* ‘live’ only survives in a few compounds, as in (40) above. By contrast, the word for ‘die’ is conservative of POc \**mate*:

- (42) \**mate* ‘die, be dead’ → HIW *met*; LTG *met*; LHI *mat*; LYP *mat*; VLW *mat*; MTP *mat*; LMG *ma?*; VRA *ma?*; VRS *miāt*; MSN *mat*; MTA *mate*; NUM *mat*; DRG *mat*; KRO *mat*; OLR *mat*; LKN *mæt*; MRL *met*

This verb \**mate* was nominalized into \**matea* (<POc \**mate-an*). Reflexes of this noun refer to the funeral ceremonies that occur after someone died, whether this refers to their burial, or to the mourning rituals that take place five, ten and fifteen days after the death:

- (43) \**matea* → ‘death; funeral ceremonies: burial, mourning rituals’: HIW *mæse*; MTP *na-mte*; LMG *mæ?æ*; VRS *mati*; MTA *matea*; DRG *mte* ~ *mata-n*; KRO *mata-n*; LKN *meta-n*; MRL *na-mati*

One of the functions of the mourning rituals is to allow the soul of the dead person to leave this world, and start its journey towards the Underworld. At this point, people no longer refer to the ‘soul’ (\**ata*) of the person, but to their ‘ghost’ (Hess 2009: 162). The etymon for ‘ghost’ reconstructs as \**atamate*. In most languages, the form has undergone apheresis (\**atamate* > \**tamate*), a phonological process common in the area; yet the Lakon form *ætmæt* makes it necessary to posit an earlier form starting in \**ata*. Following the same reasoning as earlier for its first element \*[*a*]ta- (§4.1.1), I choose to represent the etymon as \*[*a*]tamate: this reflects the fact that while some languages point to a longer form \**atamate*, others unambiguously reflect a truncated form \**tamate*.

The word is historically a compound \*[*a*]ta-*mate*, literally ‘dead person’:

- (44) \*[*a*]ta-*mate* ‘dead person, ghost+’ → HIW *təmet*; LTG *təmet*; LHI *n-təmat*; LYP *n-tamat*; VLW *n-tamat*; MTP *na-tmat*; LMG *?ama?*; VRA *?ama?*; VRS *timiāt*; MSN *tamat*; MTA *tamate*; NUM *tamat*; DRG *tmat*; KRO *tamat*; OLR *tamat*; LKN *ætmæt*; MRL *ne-temet*

While the first element \*[*a*]ta- has become opaque in modern languages, it is often easy for modern speakers to draw a connection with the word ‘dead’ (42). Note that it is ambiguous whether the first element of this compound is here (11) \*ata ‘person’, or (38) \*ata ‘soul, spirit’. The common assumption is to opt for the first interpretation, as witnessed by the POc reconstruction \*qata-mate glossed ‘dead person’ (Ross, Pawley and Osmond, in prep.). But the northern Vanuatu area leaves the ambiguity open.

#### 4.4.2 The polysemy of \*[*a*]tamate

Yet the main interest of (44) \*[*a*]ta-mate lies less in its etymology – which is mostly transparent – than in the impressive polysemy it displays in northern Vanuatu languages. While most languages of the area share the same extensions, some have gone further than others. Figure 4 represents the various meanings that can be lexified by modern reflexes of \*[*a*]tamate. It takes the form of a lexical semantic map, in the spirit of Pawley (2005) or François (2008b).

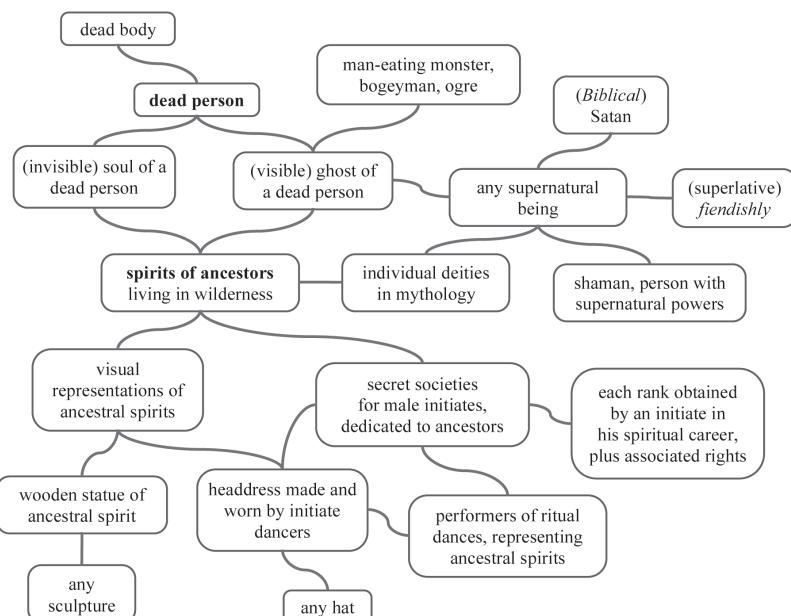


Figure 4: Lexical map showing the maximal polysemy associated with reflexes of \*[*a*]tamate ‘dead man, ghost, spirit’+

Even though \*[*a*]tamate can refer to a corpse, its most common reference is to the dead, as creatures endowed with some form of *post mortem* life. This is thus the term for ‘ghost’ – whether the phantom of a specific individual, or an unidentified spectre. The same word commonly designates any man-eating monster, ogre or bogeyman in stories told to children. Their nature of ‘living dead’ explains their hunger for fresh human flesh, whose presence they feel mostly through the sense of smell. These ghosts form a populace of nameless and indistinct spirits, who haunt the depths of the forest and of the night, always ready to harass the unsuspecting mortal. They haunt various places in the islands (§4.5.2), which are feared and avoided for that reason.

By extension, \*[*a*]tamate is sometimes the cover term for any supernatural being or spirit, whether it originates in a dead human or not. Even though it may refer to benevolent spirits – including the \*βui, the higher deities of ancestral mythology (§4.4.4) – its frequent association with scary creatures may be the reason why \*[*a*]tamate was equated, in the Christian religion, to Satan.

But the culturally most significant meaning of \*[*a*]tamate is when it refers collectively to the spirits of ancestors. These are normally thought of as benevolent towards mankind, as they take care of mortals and guide them through existence; as such, they are both feared and respected. This association ‘ghost–ancestral spirit’ is in turn the source of a cascade of metonymies that are central to customary representations. Initiation rituals, which have partly fallen into disuse today though they survive in some locations, entail that each male individual, during the course of his adult life, has to go through a number of steps in his spiritual career. Each step corresponds to a distinct circle or “secret society” (Codrington 1891: 69–92), and is associated with a specific guardian deity or ‘ancestral spirit’; for example, on Motalava, the third rank of initiation was set under the aegis of *Niwi*s, the ‘Owl’ spirit. The word \*[*a*]tamate refers to that guardian deity, but also, by metonymy, to the ‘secret society’ set under its protection. In a way similar to what happens in other “status-alteration systems” also found in Vanuatu (Bolton 2003), each of these societies constitutes a distinct rank that comes with particular rights and obligations – all subsumed under the word \*[*a*]tamate.

The activities performed by the initiates of each rank include learning about ancestors through myths, songs and visual representations of all sorts. They learn how to carve sculptures of specific spirits; in Mwotlap, the word \*[*a*]tamate not only refers to these religious icons, but can even be extended to any figurative sculpture – in competition with *nō-tötystal* we saw earlier (§3.4). Finally, an essential embodiment of the ancestral spirits is in the form of adorned head-dresses that will be worn by initiate dancers, as they perform ritual dances in public (Vienne 1996). Once again, the noun \*[*a*]tamate will refer not only to the

spirits themselves, but also to the dancers who embody them, as well as to their headdresses; by extension, the word can sometimes refer to any headdress, mask or hat (Codrington 1891: 80).

All these senses are represented in Figure 4. Languages differ as to the precise semantic extent of their reflex of \**[a]tamate*. Thus in Mota, the word *tamate* colexifies all these meanings; so does *na-tmat* in Mwotlap, except for the superlative use. In several languages though, this noun undergoes the competition of other words: thus the evil ghosts of stories are called alternatively *timiat* or *œrœr* in Vurës; *?ama?* or *ɔɔr* in Vera'a; *ætmæt* or *tanjɔ* (§4.8) in Lakon. In Mota, ritual headdresses are called either *tamate* or *kokoa* (François, pers. data). In the two Torres languages, *tømet* is restricted to positive uses (ancestral spirits; secret societies; ritual headdresses, also called *kʷəy̥ar*) – whereas evil spirits are called by other names – HIW *ŋʷe*, LTG *ŋʷie* or *βaβəro*. These local innovations are probably more recent than the general polysemy of \**[a]tamate* represented in Figure 4.

The broad polysemy of \**[a]tamate* deserves a central place in any study of etymology in this group of islands. While the literal meaning \**ata-mate* ‘dead person’ leads naturally to ‘ghost’, it takes a longer semantic path to reach such senses as ‘fiendishly’, ‘sorcerer’, ‘Satan’ or ‘hat’.

#### 4.4.3 Words and phrases based on \**[a]tamate*

The noun \**[a]tamate* is a significant etymological source not only due to its internal polysemy, but also through the many compound words and phrases which are based on it. I will only mention here a selection.

When an edible fruit has a variety that is inedible for humans, the latter is sometimes called ‘[fruit] for ghosts’. Thus in Mwotlap, the sweet *na-ŋβɪŋ* ‘Malay apple, *Syzygium malaccense*’ contrasts with the unedible variety *na-ŋβɪŋ yan tamat* (lit. ‘Malay-apple for ghosts’) ‘*Syzygium jambos*’.

The two hollows on the small of one’s back are called \**mata-i [a]tamate* ‘ghost’s eyes’:

- (45) \**mata-i [a]tamate* ‘ghost’s eyes’ → ‘hollows in small of back’: MTP *na-mte-tmat*; MTA *mate-tamate*

Not surprisingly, some compounds based on \**[a]tamate* are names of individual spirits from the mythology. Each one gives its name to the society of initiates who worship it, as well as to a ritual dance performed to celebrate it. Here are three examples:

- (46) \*[*a]tamate wasawasa* ‘smooth ghost’ → ‘name of a spirit’: MTP *na-tmat wawah*; MTA *tamate wasawasa*
- (47) \*[*a]tamate memeyeli* ‘bow-legged ghost’ → ‘name of a spirit; a manly dance’: MTP *na-tmat-mumyil*; MTA *tamat-memeyel*
- (48) \*[*a]tamate werewere* ‘whining ghost’ → ‘name of a spirit, whose whining voice is heard in the dark and scares away the profane; name of the secret society that embodies this spirit’ MTP *na-tmat-wejwej*; MTA *tamate-werewere*

When a young man is initiated to a secret society, he becomes \**lesu [a]tamate* ‘entitled to ghosts’ – that is, ‘entitled to have symbolic access to the world of ancestral spirits’:

- (49) \**lesu [a]tamate* ‘entitled to ghosts’ → ‘initiated, admitted to a society’: LTG *lh-tømet*; LYP *lis-tamat*; MTP *lis-tamat*; LMG *lis-?ama?*; MTA *les-tamate*

Only they will be allowed to ‘hold spirits’, that is, create representations of these spirits in the form of sculptures, tattoos or headdresses:

- (50) \**tauri [a]tamate* ‘hold ghosts’ → ‘create visual representations of spirits’: LHI *hv\tj n-tømat*; LYP *føj n-tamat*; VLW *tj n-tamat*; MTP *tj na-tmat*; LMG *?ør n-?ama?*; DRG *tur o tmat*; KRO *tur o tamat*; LKN (*yih*) *ætmæt*

The word \**[a]tamate* occasionally forms compounds referring to supernatural and other awe-inspiring forces of the world:

- (51) \**[a]tamate yaruyaru* ‘wading ghost’ → ‘tornado, waterspout’ MTP *na-tmat yejyej*; MTA *tamate yaruyaru*

Similar phrases are used to name man-made items whose power or effect triggers the same sort of awe as ghosts or monsters:

- (52) \**[a]tamate ti<sup>m</sup>b<sup>w</sup>a* ‘shooting ghost’ → ‘(anc.) bamboo blowpipe; (mod.) gun, rifle’: MTP *na-tmat tikp<sup>w</sup>*; VRS *timiat tikp<sup>w</sup>*; MTA *tamate tikp<sup>w</sup>a*

- (53) \*[*a*]tamate *wosa* ‘slapping ghost’ → ‘a membrane drum of northern Banks Islands, whose thick cover of leaves is punched to obtain a loud, deep sound’<sup>13</sup>

LYP *n-tamat-wɔɔs*; MTP *na-tmat-wɔh*; LMG *?ama-wɔɔs*; VRS *timiat-wɔɔs*

- (54) ‘small ghost’ → ‘harmonica, radio, tape-recorder, loudspeaker’  
HIW *wte təmet*

A last phrase involving \*[*a*]tamate is the name of the ‘shaman’ in the Torres islands, \*[*a*]tamate *roño*. It will be the object of a discussion in §4.6.

#### 4.4.4 The many kinds of spirits

Even though reflexes of \*[*a*]tamate are everywhere the cover term for most supernatural beings, each local culture has its own names for different kinds of spirits, whether good or evil, minor or prestigious.

To take just a sample from the language Mwotlap, evil phantoms include *Lisipsip* ‘a female creature with long hair, who hides in banyan roots and kidnaps children’; *Napsisyɔn* ‘a scary female ghost with long breasts, who cries on the road at night and brings a curse to childbirth’; *Nej̄mʷε* ‘Seasnake, a succubus who seduces her victims’; *Nukus* ‘a terrifying ghost’; *Jeʷbek* ‘a child-eating ghost’, and so on. In addition, Mwotlap folklore mentions small imp-like creatures who live in the bush, and are not deemed harmful: some of them are called the *Kakamɔj*<sup>14</sup> ‘elf-like creatures who live in the forest’, others are *Nrʷdrʷbit* ‘fairy-like creatures who make and give away shellmoney’; others again *Natʷbunʷbun* ‘fairy’.

While other languages occasionally have spirits with similar names, it is more often the case that each community has developed its own idiosyncratic folklore. For example, Gaua people also have their elves (François forthc.), but they are called \*Watayoriyori (>DRG *Watyvryvur*, LKN *Witay:yr:*), a word with no known etymology, and no relation with the names used in Mwotlap. It would be an interesting topic to compare the names and attributes of various creatures of the folklore in this region, and beyond.

13 See François and Stern (forthcoming).

14 Cf. the Kakamora creatures of Makira, in the Solomon Islands (Fox and Drew 1915; Fox 1962).

#### 4.4.5 The primal deities

Despite the variety of terms, one root stands out as a widespread and important word, besides \*[*a*]tamate. This term is \*βui, which may perhaps be glossed ‘primal spirit’ or ‘deity’:

- (55a) \*βui → ‘primal spirit; deity’:

HIW *wt*; LTG *wt*; LHI *n-βu*; LYP *n-βu*; VLW *n-βu*; MTP *nu-βu*; LMG *n-βu*; VRS *βü*; MSN *βu*; MTA *βui*; DRG *βu*; KRO *βu*; LKN *βu:*; MRL *nu-βu*

Codrington (1891), and after him Ivens (1931), discuss the reference of *Vui* in the “religion of Mota”; their observations are mostly confirmed by what I heard from other languages, and can safely be assigned to the protoforms. On the one hand, \*[*a*]tamate primarily refers to “ghosts”, i.e. spirits emanating from dead mortals; they are mostly seen as dangerous creatures, who would be invoked, for example in malevolent sorcery. On the other hand, \*βui designates the eternal spirits of the place, who were present even before mankind, and still inhabit the forest. These primal spirits created the world as demiurgic forces: they can legitimately be compared to deities or “gods” (see §4.4.6), whose prestige and aura rank much higher than \*[*a*]tamate. Ghosts (\*[*a*]tamate) are normally visible, and keep or take human shape to deceive their victims. By contrast, primal spirits \*βui are immortal and invisible, present in the very fabric of the land.

The etymology of \*βui is unclear. Among the ones discussed by Ivens (1931), the most promising may be POc \*puqun ‘base (of tree); root, origin’ – because these deities were present at the very beginning of Time. This tentative hypothesis is supported by the existence of similar metaphors in modern languages – e.g. Koro *l-kpʷit̚i βvnv* ‘the Origin of the World’, literally ‘at the root of land’ (François forthc.). It is possible that \*βui can further be analysed as a radical \*βu (<\*puqun) followed by the suffix \*-i (POc \*-qi) of inalienable nouns (see François 2005: 488); if so, the string \*βu-i could receive the literal gloss ‘the Origin’. Due to the brevity of the forms involved, this particular hypothesis remains speculative.

The contrast between \*βui and \*[*a*]tamate is sometimes preserved to this day. Banabas Womal, a knowledgeable man from Koro, explained to me their difference in essentially the same terms as those reported by Codrington. Hess (2009: 162) glosses *Vurës /βü/* as ‘pure spiritual being, non-human’, and contrasts it with other forms of spirits. But in some languages, the reflexes of \*βui and of \*[*a*]tamate have become almost synonyms – at least according to some speakers. Thus in Lehali, even evil spirits can be referred to using \*βui – in the phrase *n-βu sæt* ‘evil spirits’ – a sense which is elsewhere restricted to

\*[*a]tamate*. In Löyöp, one can even speak of *βu tamat* ‘evil spirits’, a phrase that includes reflexes of both roots. In two languages, the term \**βui* has taken an abstract meaning, that refers to the evil force of certain creatures – including of ghosts:

- (55b) \**βui* → ‘supernatural power of a ghost or spirit; evil force’: LTG *wu*; LMG *n-βu*

In this sense, \**βui* is taking over a meaning which is normally expressed by other roots, such as \**mana*, \**lalaβi*, \**tayaro* (§4.6).

In modern times, the same term \**βui* is also used to translate the Christian concept of Spirit (MTP *nu-βu jɔŋ* ‘the Holy Spirit’ – see §4.6) or God (HIW *Wu*).

#### 4.4.6 The pantheon of northern Vanuatu

Some \**βui* spirits are named, and constitute, as it were, the pantheon of northern Vanuatu mythologies.

The name of the main deity of the Banks Islands is known in the literature as “Qat”, after its spelling in Mota orthography. His name reconstructs as \**mBwatu*:

- (56a) \**mBwatu* → ‘name of the principal cultural hero of the Banks Is, creator of the world and of institutions’: LYP *kpwit*; VLW *ŋgbwet*; MTP *i|kpwet*; LMG *kpwet*; VRA *kpwɔ?*; VRS *kpwet*; MSN *kpwet*; MTA *kpwat*; NUM *kpwat*; DRG *kpwat*; LKN *kpwat*; MRL *kʷet*

The cycle of Qat is the most famous myth of the Banks Islands (Codrington 1891: 156; Vienne 1984; François forthc.). Qat is the deity who created the world; who brought the alternation of day and night; who sculpted the first woman; who designed some of the laws, as well as the language of songs.<sup>15</sup> Ivens (1931: 159) proposes to link his name to the noun (21) \**mβwatu* ‘head’ – a link which is possible, though not entirely convincing.

One connection that is uncontroversial is the one made by Codrington (1891: 86–92) between the name of the hero “Qat” and the *Qatu*, the name of secret societies that were in use in the islands further South (Maewo, Ambae, Pentecost). Note in passing the usage of designating these secret societies with the name of

<sup>15</sup> While Torres Islanders tell similar stories, they call the hero with another name: see below, under (60).

a spirit or god: this is parallel to the designation, in the Banks islands, of these societies with the polysemous term \*[*a]tamate* ‘ancestral spirit’ (§4.4.2). Even though the Banks languages never designate these societies with the name of their god \**mBwatu*, they still draw a connection between his name and the secret world of young initiates. Indeed, the most prestigious dance of the Banks Islands – one in which male initiates, covered with headdresses, embody ancestral spirits (Codrington 1891: 85; François 2008a, François and Stern, forthc.) – is designated by names that also point to a protoform \**mβwatu*:

- (56b) \**mβwatu* → ‘a highly prestigious dance with a secret song, performed by initiates dressed as ancestral spirits’: MTP *ne-kpwet*; VRS *kpwet*; MSN *kpwet*; MTA *kpwat*; DRG *kpwat*; LKN *kpwat*; MRL *ne-kʷet*

Qat is constantly bullied by his eleven brothers, who are all called \**Tayaro*:

- (57) \**Tayaro* → ‘a mythological figure, elder brother of the hero Qat’: VLW *Tayaj*; MTP *Tayaj*; LMG *Tayar*; VRS *Tayar*; MSN *Tayar*; MTA *Tayaro*

Among the brothers, two characters are called \**Tayaro Lolomarani* ‘smart Tagaro’ and \**Tayaro Loloimbwoni* ‘foolish Tagaro’ – see (30)–(31) above. This name \**Tayaro* is reminiscent of a major Oceanic deity, found further south in Vanuatu – under the name *Tagaro* – and more famously in Polynesia – as *Tangaroa* (Marck 1996, Biggs 2006). I will come back to this connection in §4.8 below, as I discuss a related term \**tayaro* ‘supernatural force embodied in a spirit or a stone’.

The main enemy of Qat, known in Mota as *Qasavara*, reconstructs as \**mBwasaβara*:

- (58) \**mBwasaβara* → ‘a giant ogre, enemy of the hero Qat’: MTP *kpwasβaj*; VRS (*kpwastaβaβ*); MSN (*wɔ|kpwastaβaβ*); MTA *kpwasaβara*; LKN *kpwasβal*

The first element of his name is \**mβwasa* ‘bald head’; the second element varies, and is opaque.

Another of Qat’s foes was evidently called \**nDule*. His name is proverbially associated with lie and deception:

- (59a) \**nDule* → ‘name of a trickster spirit, who constantly deceives Qat’: MTP *i|ndvl*; LMG *tvl*; VRS *nvl*; MSN *nvl*; MTA *nule*

Stories of *\*nDule* are plentiful on Vanua Lava (Hess 2009: 149, 219), but practically unknown elsewhere. Interestingly, knowledge of regular sound correspondences make it possible to propose a connection between the name of this deity and a word used in the Torres Islands to designate a magic stone:

- (59b) *\*ndule* → ‘a standing stone endowed with supernatural powers, inhabited by a spirit; a stone statue representing such a spirit’:  
 HIW *tūjə*; LTG *tūla*

Torres islanders used to erect these magic stones at the entrance of their *\*yamali*, the sacred houses of high chiefs (§4.7.1). They still fear them nowadays.

Another rival of Qat is *\*Marawa* ‘Spider’:

- (60) *\*marawa* ‘spider’ → *\*Marawa* ‘name of a spirit who first deceives Qat, but eventually becomes his ally’:   
 HIW *məgLawə*; LTG *mərawə*; LHI *n-majaw*; LYP *n-majaw*; VLW *n-majaw*; MTP *na-mjaw*; LMG *n-morow*; VRA *marawa*; VRS *maraw*; MSN *maraw*; MTA *marawa*; DRG *mraw*; KRO *maraw*; OLR *maraw*; LKN *maraw* ~ *wimarawraw*

This mythological figure is of secondary importance in northern Banks, and only present as a character in the myth of Qat. However, in the Torres Is, the main hero of myths – often equated with the Qat of the Banks Is – is called HIW *məgLaftit*; LTG *mərawəhīh*, two forms which may be related to *\*Marawa*.

The name *\*marawa* also appears to bear some significance on Gaua (François forthc.). Thus in Lakon, *maraw* designates a type of ritual headdress that is worn by initiated young men, along with the more noble *ætmæt*. Besides, all languages of Gaua (except Nume) designate their traditional myths and stories as literally ‘Stories of Spider’:

- (61) *\*susur[ay]i m<sup>b</sup>ula-i Marawa* ‘stories of Spider’ → ‘traditional story, myth’:   
 DRG *susriy<sup>m</sup>ble-maraw*; KRO *ususreay<sup>m</sup>bele-maraw*;   
 OLR *ususra-puli-maraw*; LKN *susu-pule-maraw*

The phrases in (61) are reminiscent of the way in which three languages of Vanua Lava call their myths, literally ‘the Voice of *βui* spirits’:

- (62) *\*aliŋa-i βui* ‘voice of higher spirits’ → ‘traditional story, myth’:   
 VRA *nliŋo-βui*; VRS *elŋe-βui*; MSN *ɔŋe-βui*

A similar example is the way northern Vanuatu people designate the special speech register they use in sung poetry (François 2008a; François and Stern

forthc.). This archaic language, perceived as an echo of primitive times and extinct voices of the past, is known in Mwotlap as *na-βap nɔn ikp<sup>w</sup>et* ‘the language of Qat’, and in Lo-Toga as *na βəyəβəyə mi həŋ<sup>w</sup>ərə təmet* ‘the language of Ancestral Spirits’.

## 4.5 The dwelling places of ancestors

The abode of the dead receives many names in the languages of northern Vanuatu. This diversity reflects not only the tendency for lexical innovation (§2.3), but also the variety of locations that can be associated with ghosts and spirits. Essentially, one can think of four different places which may be considered the abode of the dead:

- the graveyard, on the side of the village
- various rocks and places in the bush, which are said to be haunted
- the Other World, or “Hades”: a specific location where spirits dwell
- the ritual location where the societies of spirits gather

### 4.5.1 The graveyard

Several languages have a phrase ‘enclosure of ghosts’ to designate a cemetery or graveyard, on the side of the village. While the original term for ‘enclosure’ (*\*ara* < POc *\*qaRa*) has been replaced in some languages, the second element always reflects *\*[a]tamate*:

- (63) *\*ara [a]tamate* ‘enclosure of ghosts’ → ‘graveyard’:   
 HIW *pajə təmet*; LTG *pialə təmet*; LHI *n-aj təmat*; MTP *ne-yeaj tamat*;   
 MTA *yeara tamate*

Due to the fear of the dead, the languages of southern Banks avoid the use of *\*[a]tamate* here, and use a euphemistic phrase ‘sleeping enclosure’:

- (64) ‘enclosure for sleeping’ → ‘graveyard’:   
 NUM *yeare nɔŋɔŋɔr*; DRG *ar-matur*; KRO *ar-matur*; LKN *peret nɔ:ŋɔ:*;   
 MRL *ne-yeare mətər*

Mwotlap has a special word for ‘graveyard’, namely *ne-tekp<sup>w</sup>* – see (67b).

Finally, a noun whose meaning is related to this domain is *\*m<sup>b</sup>waru* ‘grave’. When the form is a suffixable noun, I cite it here with a 3sg possessor *\*-na* (‘his/her grave’):

- (65a) \**m̥bw̥aru* (-na) → ‘grave (of s.o.), dug in the ground’  
LTG *kʷərū-nə*; VLW *n-ŋgbʷvju-n*; MTP *ni-kpʷjv-n*; VRS *kpʷœr*;  
NUM *na-kpʷurv-n*; DRG *na-kpʷvrkpʷurv-n*; KRO *na-kpʷvrkpʷurv-n*;  
LKN *kpʷarv-n*

In the two Torres languages, the same etymon \**m̥bw̥aru* has become the name of a special type of grave found only in these islands – namely, a large stone tomb or mausoleum which was erected only for highly ranked chiefs:

- (65b) \**m̥bw̥aru* → ‘stone tomb for highly ranked chiefs’:   
HIW *kʷɔgₗ*; LTG *kʷɔr*

In the two Torres languages, the meaning ‘grave in the ground’ is not lexified by \**m̥bw̥aru* any more: it has been replaced by a noun *ŋwet*, which originally designates the earth mounds created as one plants yams in the garden. Likewise, Mota calls the grave *tawe tamate*, literally ‘a dead man’s hill (mound)’.

The original meaning of the etymon \**m̥bw̥aru* can be reconstructed as ‘individual grave (of s.o.)’. However, in two languages, it has undergone semantic shift to refer to the whole area where graves are grouped together, namely ‘graveyard’:

- (65c) \**m̥bw̥aru* → ‘graveyard’  
VLW *nε-ŋgbʷɛj*; MRL *nɔ-kʷɔr*

We’ll see below a final example of semantic shift affecting \**m̥bw̥aru* in the language Löyöp – see (68).

#### 4.5.2 Haunted places in the island

Aside from the cemetery, another place where ghosts are to be found is in their various dwelling places across the islands – typically on rocky capes and promontories. These places receive various names; a common one is simply ‘bad places’ (e.g. LKN *l̥ ya sa*), because these are locations where people fear to catch diseases or suffer from other misfortunes. Another simple name is ‘place of ghosts’ – e.g. VLW *n-tɔnɔ tamat*.

Hiw and Lehali call these locations ‘taboo places’:

- (66) ‘taboo places’ → ‘places on the island haunted by ghosts’:   
HIW *nə ŋwutə tɔkʷ*; LHI *nε-tpu*

These two phrases make use of an adjective ‘taboo, sacred’ which conveys both awe and fear. The word has reflexes with the same meaning in a few languages of the area:

- (67a) \**taʷbu* → ‘taboo, holy, sacred; worthy of awe and fear; off limits, unapproachable’:   
HIW *tɔkʷ*; LTG *tɔkʷ*; LHI *tpu*; MTA *tap ~ tapu*; DRG *taʷb*; KRO *tēam*; MRL *tɔm*

The word /*tɔkʷ*/ reflects a form with a labio-velar consonant \**taŋgbʷu*, itself a regular reflex of \**taʷbu*.<sup>16</sup> The latter results from the rounding before /u/ of \**mb*; but the unrounded variant was preserved in the other languages further south. The etymon to be reconstructed here is thus ultimately \**taʷbu* ‘sacred; forbidden’ – that is, the very same word which, via Polynesian *tapu*, was borrowed into English as *taboo*.

Keesing (1985: 204) suggests that the best gloss for this common Oceanic word should be ‘off limits’. He also points out the inherently relational nature of the term: “Something that is off limits, *tapu*, is always off limits to *someone*, not in and of itself; (...) being *tapu*, off limits, implies a context.” This observation helps us understand the different meanings that are associated with this root \**taʷbu* in our Torres–Banks languages. Thus Mwotlap has a noun *ne-tekpʷ* meaning ‘cemetery’, the regular reflex of \**taʷbu* < \**taʷbu*:

- (67b) \**taʷbu* → [N] ‘cemetery, graveyard’: MTP *ne-tekpʷ*

A cemetery is “off limits” to humans, because it is the dwelling place of the dead \*[a]tamate; it is an area which all villagers, in normal circumstances, will want to avoid.

In (66) and (67b), the focus was placed on a location associated with the dead (haunted rocks, cemetery), and the symbolic interdiction was directed at the living. Yet sometimes the perspective is reversed. Thus in Mwerlap, the noun *nɔ-tɔm* refers to a conventional sign made with certain leaves, that is put up at an entrance in order to deny access to undesirable guests [cf. (76)]; it is especially used to scare evil spirits away from a newborn baby (Marie Durand, p.c.):

<sup>16</sup> Similar correspondences between a reconstructable \**m̥bw̥* and a rounded velar *kʷ* in Torres languages, can be seen in (21); (65b); (76b); (80). The vowel is also regular, as shown in the rule in (6).

- (67c) \**ta<sup>m</sup>bu* → [N] ‘a sign marking a place as taboo or forbidden, especially against ghosts’: MRL *nɔ-tɔm*

In this particular case, the focus of \**ta<sup>m</sup>bu* is on a location associated with the living (the house of a newborn baby), and the taboo sign is directed against the ghosts.

Finally, besides its adjective *tap* ‘taboo’, Mota has a noun *tap* that refers to a day of mourning ordered by a secret society (*tamate*), when no drumming, singing or playing are allowed (Codrington and Palmer 1896: 196). This obviously has the same origin:

- (67d) \**ta<sup>m</sup>bu* → [N] ‘day of silence in sign of mourning’: MTA *tap*

This sense too illustrates the strong connection that exists between the word \**ta<sup>m</sup>bu* and the fear of death. The same root \**ta<sup>m</sup>bu* will be mentioned again in (76), and in §4.6.

Löyöp draws a connection between ‘grave’ or ‘graveyard’ on the one hand, and ‘place haunted by spirits’ on the other. It lexifies the latter meaning with a regular reflex of the root \**mbw<sup>w</sup>aru*, which we saw in (65a) above with the sense ‘grave’:

- (68) \**mbw<sup>w</sup>aru* ‘grave’ → ‘area in island haunted by ghosts and spirits’: LYP *nø-kp<sup>w</sup>øj*

Ancestral spirits are known for their love of dancing: many stories portray them as a group of merry ghosts who spend their nights in frenetic dances, somewhere in the bush. It is therefore not an accident that the noun \**sara*, which normally designates the open space in the midst of villages where dances are performed (Clark 2009: 238; François 2011a: 183), sometimes designates the dancing place of ghosts in the wilderness. This is done through a phrase which can be reconstructed as \**sara-i* [a]*tamate*:

- (69) \**sara-i* [a]*tamate* ‘dancing-place of ghosts’  
→ ‘the abode of ancestral spirits, in the wilderness’: LYP *l̩ s̩i<sup>j</sup>-tmat*; MTP *le-hjε-tmat*; LMG *sere-ʔama?*; VRA *sere-ʔama?*; VRS *sere-timiat*

This word has given its name to the highest volcano on Vanua Lava island, shown on maps as Suretimeat or Sere’ama. It is considered by the islanders to be the meeting place of spirits for their dances (Hess 2009: 166).

#### 4.5.3 The abode of the dead

Most of the words we have seen in §4.5.2 are common nouns: they can refer to more than one place in the same island. By contrast, it is not rare to hear the idea that ghosts and spirits, even though they occasionally inhabit various places in the wilderness, really belong to one specific place where they dwell for eternity – what we call Hell, the Underworld or Hades. Interestingly enough, the representations of Hell have been little influenced by the otherwise widespread Christian religion. Far from coinciding with dual representations of “Paradise” or “Hell”, the terms I was able to collect for the Underworld – either through interviews or through myths and stories – all reflect an ancient system of representations that remains vibrant today. They are also consistent with other accounts by previous observers (Codrington 1891; Vienne 1984).

A first cognate set involves a root \**sura* (or rather, as we’ll see below, \**asura*), preceded by a prefix of unclear origin:

- (70a) \*[a]*sura* → ‘abode of the dead, Hell’: LYP *wv\svj*; MTP *wv\svj*; LMG *wirr\sur*; VRS *wirr\sur*

The people of Gaua use the same term for the abode of the dead, but interestingly, they are very specific as to its location. Their myths explicitly locate the Other World under their majestic Gharet volcano (François forthc.):

- (70b) \*[a]*sura* → ‘abode of the dead, located under a volcano’: DRG *wrr\sur*; KRO *wirr\sur*; OLR *wirr\svj*; LKN *wore\hv*

Thus, a Dorig myth explains how Wrisris, the god of Death and guardian of the Underworld, has his dwelling under the volcano of Gaua.

Interestingly, the Lakon word *ahv*: ‘volcano’ (in which the long vowel reflects \*r, and /h/ < \*s) is cognate with these sets. The tendency of Lakon to preserve pretonic vowels (François 2005: 470) requires the reconstruction of an initial vowel \*a to the etymon, hence \**asura*:

- (70c) \**asura* → ‘volcano’: LKN *ahv*

This Lakon name for ‘volcano’ seems to be a local innovation; it is distinct from the other root which can be reconstructed for this meaning, namely \**buro*:

- (71) \**buro* ‘volcano’ → ‘volcano’: LHI *n-βɔj*; LYP *n-βvj*; VLW *nu-βvj*; MTP *nv-βvj*; LMG *n-βvr*; VRS *βvr*; MSN *βur*; MTA *βro*; NUM *wa\ur*; DRG *vr*; OLR *n\oj*

In other words, the original word for ‘volcano’, in the northern Vanuatu area, is *\*βuro*, which bears no lexical relationship with the world of the dead. As for *\*asura*, the origin of Lakon *ahvə*, it must have had a more specific meaning. Codrington and Palmer (1896) has a cognate form in Mota, with an interesting gloss:

- (70d) \*[a]sura → ‘narrow entrance; esp. entrance into Hades’: MTA *sura*

Likewise, regular correspondences make it clear that the following forms in the two Torres languages – despite their profane meaning – stem from the same word:

- (70e) \*[a]sura → ‘deep and narrow cave’: HIW *sugLə*; LTG *hurə*

All these forms suggest to reconstruct a word \*[a]sura, defined as ‘deep and narrow opening inside a mountain or volcano, leading to the Underworld’.

But if the original meaning of \*[a]sura was only ‘narrow entrance towards the Underworld’, then what was the name of that Underworld? In fact, these northern Vanuatu languages have a second term to designate the abode of the dead, *\*mbanoi*:

- (72a) *\*mbanoi* → ‘Hades, the abode of the dead’: HIW *pene*; LTG *pənə*; LHI *pni*; LYP *pənə*; VLW *a|mnu*; MTP *a|mnu*; VRS *mbønø*; MTA *panoi*; LKN *pani*

Hess (2009: 166) reports on interviews she had with a speaker of Vurës. On the one hand, he interpreted *Wirisvə* (70a) as a “place of sorrow and sadness where souls receive their punishment for their immoral and evil practices”, and compared it to the Christian “Hell”. On the other hand, *kpwaraŋ mbønø* (in which *kpwaraŋ* means ‘cave’) was equated with “Paradise”, and described as a place of “joy and happiness”. I have heard nowhere else of such a dichotomy between two realms for the dead that so closely reflects syncretism with Christian beliefs.

Interestingly, this Vanua Lava islander located Paradise under the volcano of the neighbouring island of Gaua. In doing so, he did not link the volcano to the etymon *\*asura* (as in *Wirisvə*), but to the other etymon *\*mbanoi* (>*mbønø*). But why? No language of the Torres–Banks area seems to associate *\*mbanoi* specifically with volcanoes. In fact, this connection is found in other languages of Vanuatu further south (Clark 2009: 80):

- (72b) *\*mbanoi* → ‘volcano’: Paama *vanei*; Namakir *mbane*; Nguna *na-panoi*

- (72c) *\*mbanoi* → ‘volcanic ash’: Tamambo *mbanoi*; Uriquiv *mbenu*; Lewo *pani*

Clearly, ancient cultures of Vanuatu have long associated the abode of the dead with volcanoes. Should one observe these words on a synchronic basis in each language taken separately, this connection would appear nowhere; it only comes to light thanks to language comparison. In sum, I propose to reconstruct the two following meanings for these two words:

- *\*mbanoi* ‘volcano; esp. the Volcano where the dead abide, the Underworld’;
- *\*asura* ‘a long and narrow cavity in a mountain or volcano; esp. the long and narrow corridor leading to the Underworld (*\*mbanoi*)’

Nowadays, the association of *\*mbanoi* with volcanoes has been mostly lost. In Hiw, the word *Pene!* has become a respectful salute when parting with someone: one says ‘[See you in] the Other World!’ to make sure that, should anything bad happen to one’s friend, their soul will safely reach the abode of the dead, rather than wander around the world like a lost soul.

Modern reflexes of *\*mbanoi* refer to an abstract location – Hell – which most modern speakers view as a remote place, often with no further specification. Except for Gaua island where the Gharet volcano is still a major landmark, elsewhere the abode of the dead is seldom identified in the actual geography of the islands; it is usually described as a ‘very remote’ place – to the point that its name is sometimes used, jokingly, to mean ‘somewhere very far’. Some speakers of Mwotlap told me that *Amnu* may be somewhere in the ocean – in conformity with the traditional link, heard in some stories, between ghosts and the sea.

The most reliable information I collected was in an interview with Sesil Pilaglique, a highly knowledgeable shaman (§4.6) of Toga island who is still active, and has travelled oftentimes to the Other World. According to him, *Pənə* (<*\*mbanoi*) is located everywhere in the island, and nowhere in particular: ghosts and ancestral spirits inhabit the same world as ours, yet are invisible. The power of shamans and seers like himself precisely consists in being able to see the phantoms around us, and interact with them. After their mystical journey, they come back to our world, and become one of us again.

By contrast with *\*mbanoi*, the World of the Living is called *\*marama*:

- (73) *\*marama* → ‘world; especially the World of the Living’:

- HIW *məgLamə*; LTG *maremə*; LHI *n-məjam*; LYP *n-majam*; VLW *n-majam*; MTP *na-mjam*; LMG *n-mvrvm*; VRA *marama*; VRS *maram*; MSN *maram*; MTA *marama*; NUM *maram*; DRG *mram*; KRO *maram*; OLR *maram*; LKN *maram*; MRL *na-maram*

The noun *\*marama* is etymologically linked with the notion of light: whether as daylight, as moonlight (→ *marama* ‘moon’ in various Oceanic languages: Ross,

Pawley, and Osmond 2003: 159) or as metaphorical light (→ *maʔama* ‘intelligence’ in Marquesan: Charpentier and François forthc.; *mārama* ‘understand’ in Māori: Pawley, p.c.).<sup>17</sup> This contrasts with the typical association of the dead with darkness – either the one that reigns in the Underworld, or the darkness of Night, when ghosts come out to dance.

#### 4.5.4 The secret societies

Finally, the last places where the dead can be said to dwell are the secret locations where they are symbolically revived by initiate societies. As an institution, these societies or clubs are usually called \*[*a*]tamate, using the same name, metonymically, as the ancestral spirits themselves – see Figure 4 in §4.4.2. However, other words are also connected to that lexical domain.

One central term refers to the secret meeting place of male initiates: \**salayoro* (Codrington 1891: 69–86; Vienne 1984: 319).

- (74a) \**salayoro* → ‘sacred enclosure in the bush where initiation rituals are carried out, and where spirit dances are prepared’: LTG *holøyor*; LHI *n-halyej*; LYP *n-salyɔj*; VLW *n-halyɔj*; MTP *na-halyɔj*; LMG *n-salyøer*; VRA *salyvr*; VRS *salyvr*; MSN *salyɔr*; MTA *salayoro*; NUM *salɔwɔr*; DRG *salyɔr*; KRO *salyɔr*; OLR *salwoj*; LKN *salwo:*

The \**salayoro* takes the form of a temporary shelter erected somewhere in the bush, at a distance from the village. It will be used only for a few weeks, during a cycle of initiation rituals; then it will be destroyed or abandoned until the next cycle, which may not take place for another few years, usually triggered by the preparation of a dance ceremony or festival. This enclosure is devoted to the education of young male initiates (49), who are to live in the bush during several weeks, away from the profane village. This is where they will have to endure physical ordeals, but also learn from their elders some secrets on a variety of subjects – on the meaning of rites and symbols, on the cosmogony and the mythology, on the ritual arts of singing, dancing or representing ancestral spirits (50), etc. These mysteries are to be kept away from the profane – women, children, foreigners. And indeed, some languages use the same word \**salayoro* to mean ‘secret, mystery’:

<sup>17</sup> For northern Vanuatu examples of the semantic link between ‘light’ and ‘intelligence’, see §4.2.

- (74b) \**salayoro* → ‘secret, mystery; taboo subject’: VLW *n-halyɔj*; MTP *na-halyɔj*; VRS *salyvr*

The precise etymology of \**salayoro* is not entirely clear. It is most probably a compound \**sala-yoro*, where \**sala* may be the noun ‘road, path, way’ (<POc \**jalan*). As for \**yoro*, it is reminiscent of an adverb \**yoro* with a highly polysemous meaning:

- (75) \**yoro* → ‘[do s.th.] so as to surround, cover, obstruct, prevent, protect...’: HIW *gloy*; LTG *yor*; LHI *yɛj*; LYP *çyj*; VLW *çy*; MTP *çy*; LMG *yær*; VRA *yur*; VRS *yur*; MSN *yɔr*; MTA *yoro*; NUM *çwɔr*; DRG *çy*; KRO *çy*; OLR *çwɔr*; LKN *tuçwɔ:*; MRL *çwɔr*

While the principal use of \**yoro* is as an adverb or verb modifier (cf. François 2000b), the word is more rarely found as an adjective, meaning ‘closed’ (cf. sense ‘obstruct’) or ‘taboo, secret’ (cf. sense ‘prevent’). If indeed \**sala* is the noun ‘path’, then my interpretation of \**salayoro* would therefore be ‘secret path’.

It is not an accident that some languages designate the same secret enclosure precisely using the root \**ta<sup>m</sup>bu* ‘taboo, holy’. We saw in (67) above that this word was often associated with the dead, via such meanings as ‘graveyard’, ‘mourning’, or ‘place haunted by ghosts’. This radical \**ta<sup>m</sup>bu* was the source of a derived verb \**ta<sup>m</sup>bu-a*, ‘render (s.th.) taboo’. In some Banks languages, it is still used as a transitive verb:

- (76a) \**ta<sup>m</sup>bu-a* ‘render (s.th.) taboo’ → ‘forbid access to one’s private territory (garden, house...), by standing up a conventional sign designed for that function’: MTP *twkp<sup>w</sup>u*; MTA *tapua ~ tapuy*; MRL *tɔm/yɔr*

In the two Torres languages, the same form has become a noun referring to the secret enclosure which non-initiates must avoid:

- (76b) \**ta<sup>m</sup>bu-a* ‘render (s.th.) taboo’ → [N] ‘sacred enclosure in the bush where initiation rituals are carried out, and where spirit dances are prepared’: HIW *tek<sup>w</sup>θ*; LTG *tæk<sup>w</sup>θ*

In Hiw, the enclosure is sometimes named just *tek<sup>w</sup>θ*, and sometimes *tek<sup>w</sup>θ tɔk<sup>w</sup>* – literally ‘a sacred-enclosure that is taboo’ – a phrase in which the etymon \**ta<sup>m</sup>bu* appears twice. In addition, Hiw has a phrase *βen tæk<sup>w</sup>tɔk<sup>w</sup>* (literally ‘to go taboo’), a cover term for the various restrictions – mostly related to food or body – undergone by boys during their initiation.

A whole set of lexemes can be found related to these initiation rituals: nouns referring to the fence of the enclosure, to symbolic divisions within it, to symbols used during initiation, to the various steps and tests taken by the candidates, and so on. However, for reasons of length, I shall refrain from discussing these.

#### 4.6 The shaman and the sacred

A culturally important concept in the Torres languages is \*[*a*]tamate *roño*, which refers to a healer or shaman:

- (77) \*[*a*]tamate *roño* → ‘magic healer, shaman’  
 HIW *tamet-<sup>g</sup>Loŋ*, LTG *tamet-roŋ*

This is the name given to a healer with shamanic powers – someone who knows how to circulate across worlds, and access the hidden universe of ancestral spirits (cf. §4.3.2 above). Shamans are seers: they can see what is invisible to us, ordinary mortals. This skill gives them the ability, for example, to meet the soul of a recently dead person, enquire about the cause of their death, and bring back the answer to the mourning relatives. But the shaman’s main role is to be a healer. When somebody is sick, this means their soul has been kidnapped by spirits (\*[*a*]tamate), and carried away to the Other World (\*<sup>m</sup>*banoi*). Only a shaman has the power – aided by magic leaves – to migrate to that world, retrieve the lost soul of the person, and lead it back to the world of the Living. Many tales tell the story of how an old sorceress was thus able to retrieve the spirit of a dead hero and bring him back to life.

The etymology of \*[*a*]tamate *roño* is tricky. The first element \*[*a*]tamate must not be taken here in its sense ‘ghost’ or ‘ancestral spirit’, and instead takes on a rarer meaning ‘supernatural being’, i.e. ‘person endowed with magical powers’ (see Figure 4 in §4.4.2). Besides, the second element \**roño* is also ambiguous. When questioned about the words in (77), modern speakers interpret them as ‘spirit who listens’. This is due to the fact that the two languages have a verb ‘hear, listen, feel’ (HIW *gLoŋ*, LTG *roŋ*) that comes indeed from a root \**roño* (POc \**roñoR*). If this is correct, then \*[*a*]tamate *roño* (lit. ‘ghost hear’) would mean literally ‘a supernatural being who listens’ – that is, a ‘shaman’.

However, I would suggest that this modern interpretation is a case of folk etymology. An alternative hypothesis would be to understand \*[*a*]tamate *roño* as meaning ‘a supernatural being (\*[*a*]tamate) that is sacred (\**roño*)’. Indeed, besides the verb ‘hear’, a homophonous etymon \**roño* also reconstructs with a

meaning ‘sacred, holy’. It has been lost in the Torres – and is therefore inaccessible to modern speakers – but preserved in some of the neighbouring Banks languages:

- (78) \**roño* ‘sacred, holy’ → LHI *jeŋ*, MTP *jɔŋ*, LMG *rœŋ* VRS *rʊŋ*, MTA *roño*, MRL *rɔŋ*

It seems that this word \**roño* may in turn be semantically related with \**roroño* ‘quiet, silent’:

- (79) \**roroño* ‘quiet, silent, peaceful’ → HIW *βa<sup>g</sup>Loŋ*, LTG *rərɔŋ*, LHI *ejjɛŋ*, LYP *jɔŋ*, MTP *jɔŋɔŋ*, LMG *rœrœŋ*, VRA *rurʊŋ*, VRS *rurʊŋ*, MSN *rɔŋɔŋ*, MTA *roroño*, DRG *rrɔŋ*, LKN *rɔŋɔŋ*, MRL *rɔŋɔŋ*

The sort of holiness evoked by (78) \**roño* is indeed one endowed with a peaceful aura of solemnity. It refers to the inherent sanctity of a person, a place or an object. While \**roño* entails a sentiment of awe and respect, it doesn’t have the negative connotation of the word we saw earlier for ‘sacred’, namely (67) \**ta<sup>m</sup>bu* ‘taboo, holy, sacred; worthy of awe and fear; off limits, unapproachable’.

Mota is one of the few languages which has preserved reflexes of both etyma, in the form of *roño* and *tapu*. Codrington and Palmer (1896) draw an explicit contrast between *roño* ‘sacred, unapproachable, with inherent sanctity, not *tapu*’ and *tapu* ‘taboo, unapproachable, not to be touched, under a prohibition with the sanction of some *mana* belonging to men; so distinct from *roño*’ (cf. Codrington 1891: 77, 181, 215). The opposition between the two types of sanctity is remarkably parallel to the contrast drawn by Benveniste (1969: 179) between two concepts of sanctity found in ancient Indo-European languages: respectively a “positive” notion – “that which is endowed with a divine presence” (Lat. *sacer*) – and a “negative” one – “that which is forbidden” (Lat. *sanctus*). The two concepts of Proto Torres–Banks seem to define a similar dichotomy – that between a positive and inherent holiness of spiritual beings (\**roño*), and the negative kind of sanctity that is sanctioned by society (\**ta<sup>m</sup>bu*).

The word \**roño* has received a Christian meaning in certain languages (Lehali, Mwotlap, Mwerlap), e.g. LHI *n-εŋ jeŋ* = MTP *n-ɪŋm<sup>w</sup> jɔŋ* ‘house holy’ ‘church’; MTP *nu-βu jɔŋ* (spirit holy) ‘the Holy Spirit’; MTP *ni-<sup>m</sup>bi jɔŋ* (water holy) ‘holy water’; MTP *βasuw-jɔŋ* (bathe holy) ‘baptize’. However, even today one can still hear phrases in which the character of holiness is linked to pre-Christian beliefs: thus a stone will be ‘holy’ (MTP *ne-βet jɔŋ*) if it is haunted by the spirits of the land (*nu-βu*, cf. §4.4.5) – cf. Codrington (1891: 181).

In sum, a plausible interpretation of the shaman's name (77) in the Torres Islands would be 'a supernatural being (*\*[a]tamate*) that is sacred (*\*roño*', i.e. endowed with spiritual powers (*\*mana*, cf. §4.8.1).

## 4.7 Power, wealth and aura

### 4.7.1 The system of chiefly grades

The societies where initiates are taught the secrets of ancestral spirits have a public, open counterpart – namely, a system of hierarchical ranks in which powerful men accessed various levels of political prestige, by going through a number of ceremonies along their lives. The rank system – which mostly disappeared from the region during the 20th century – is described in the literature (Codrington 1891; Vienne 1972, 1984) as "suqe", after its name in the orthography of Mota. From the linguist's perspective, the word's etymon reconstructs in Proto Torres–Banks as *\*su<sup>m</sup>b<sup>w</sup>e*:

- (80) *\*su<sup>m</sup>b<sup>w</sup>e* → 'political system based on a hierarchy of ranks, which individual men achieved through display of riches':  
 HIW *suk<sup>w</sup>ə*; LTG *huk<sup>w</sup>ə*; LHI *n-sɔk*; LYP *n-svŋ*; VLW *n-svŋj̪m<sup>w</sup>*; MTP *nu-svkp<sup>w</sup>*;  
 LMG *n-sv̪k*; VRA *sukp<sup>w</sup>ɔ*; VRS *sukp<sup>w</sup>*; MSN *sukp<sup>w</sup>*; MTA *sukp<sup>w</sup>e ~ skp<sup>w</sup>e*;  
 DRG *β|svk*; KRO *βv|svk*; LKN *βa|svk*

Reflexes of *\*su<sup>m</sup>b<sup>w</sup>e* are used as a noun but also as a verb, meaning 'take the steps to enhance one's grade in the rank system'. This term does not appear to be a compound, and has no clear etymology. However, it is useful to note that Vanuatu languages further south, which had a different political system, have words that are cognate with *\*su<sup>m</sup>b<sup>w</sup>e*: e.g. Araki *supe* 'supreme chief, high dignitary; the Christian God' (François 2002: 302); Nguni *na-sup<sup>w</sup>e* 'idol, image of ancestor, god' (Clark 2009: 183).

The chiefly members of a *\*su<sup>m</sup>b<sup>w</sup>e* would meet in a special house reserved for them, the *\*yamali* (<POc *\*kamaliR* 'men's house'). The *\*yamali* was a house where highly ranked men would – among other activities – consume kava,<sup>18</sup> a prestigious beverage made from a plant (*Piper methysticum*), which was often associated with magic practices. With the Christianisation and the disappearance of *\*su<sup>m</sup>b<sup>w</sup>e* practices during the 20th century, both the beverage and the men's

<sup>18</sup> In this part of Vanuatu, the etymon for kava apparently reconstructs as *\*yaya* (François 2005: 478), and ultimately originates in a POc etymon *\*kawaR* 'root' (Lynch 2002).

house have largely lost their sacred aura: they have become part of the profane daily life of modern islanders.

The way for a man to climb the political scale of *\*su<sup>m</sup>b<sup>w</sup>e* involved the public display of considerable wealth. This would take the form, typically, of a number of pigs (*\*mb<sup>w</sup>oe*) that were killed by being struck (*\*βusi*) on the head with a wooden club (*\*mb<sup>w</sup>oroni*). Besides, the candidate had to bring offerings of kava, along with massive quantities of shellmoney (*\*somu*). This shellmoney consisted of small cone shells (*Conus* sp.) that had been patiently filed into circular discs, then pierced and threaded onto a very long string of beads:

- (81) *\*somu* 'Conus shell'  
 → 'traditional shell money, made of cone shells; (hence) modern money':  
 LHI *n-sim*; LYP *n-søm*; VLW *n-sim*; MTP *ni-sim*; LMG *n-søm*; VRS *søm*; MSN *sum*;  
 MTA *som*; NUM *sum*; DRG *sum*; OLR *sim*; LKN *him*; MRL *ne-søm*

The quantity of such money required for some higher ranks could measure up to 10 fathoms in length, and involved considerable work on part of the candidate's female relatives. This shellmoney has now disappeared from most places, and only exists vestigially in some families; however, it is still in currency in Merelava (Marie Durand, p.c.) shellmoney used to be a marker of one's wealth. The equivalent of Eng. 'rich' is a phrase which can be reconstructed as *\*m<sup>w</sup>era-i somu*, literally 'child of shellmoney' [cf. (17)]:

- (82) *\*m<sup>w</sup>era-i somu* 'child of shellmoney' → 'rich, wealthy':  
 LHI *n-η<sup>w</sup>tji-sim*; LYP *η̄m<sup>w</sup>tji-søm*; MTP *na-η̄m<sup>w</sup>je-sim*; VRS *η̄m<sup>w</sup>ørø-søm*;  
 MSN *η̄m<sup>w</sup>ere-sum*; MTA *η̄m<sup>w</sup>ere-søm*; DRG *η̄m<sup>w</sup>re-sum*; LKN *η̄m<sup>w</sup>ri-him*

### 4.7.2 The aura of high chiefs

While the detail differed from island to island, the usual number of political ranks on the scale was between 12 and 14. Each rank had its own attributes (Vienne 1984), and its own name. I will only cite three of them here:

- (83) *\*la<sup>n</sup>do* → 'name of a chiefly rank, relatively high':  
 LYP *wv|lan*; VLW *na-lan*, *we|lan*; MTP *na-lan*, *we|lan*; MTA *lano*; LKN *lat̪*  
 (84) *\*wo mata-i aloa* (lit. 'the eye of the sun') 'rising sun, Orient'  
 → 'name of a chiefly rank, among the highest':  
 VLW *wɔmtelɔ*; MTP *wɔmtelɔ*; VRA *wɔmitilɔ*; VRS *wemtelɔ*; MSN *wɔmetelɔ*;  
 MTA *wometeloa*; DRG *wamtalɔ*; LKN *wimat<sup>2</sup>alɔ*; MRL *wemtelɔ*

- (85) \**tayuru-i maliyo* ‘behind the clouds’  
 → ‘name of a chiefly rank, among the highest’: LTG (*kile tukə*); VRS *tewrū maly*; MTA *we taur o maliyo*; DRG *tawri mly*; LKN *tawu: maly*; MRL *tewur mely*

The etymologies of the last two sets show how the prestige of high chiefs tends to be expressed with celestial metaphors: ‘Rising Sun’, ‘Beyond the Clouds’. I also include here the Lo-Toga form *kile tukə*, even though it is not cognate with the rest, because the name of this high rank uses a similar metaphor: ‘beyond the firmament’.

A man having reached one of the higher ranks would have gained considerable social prestige. He would then be called \**taβusimwele*:

- (86) \**taβusimwele* → ‘man of great rank, high chief’: MTP *na-tbusiŋmʷel*; VRA *tuβusiŋmʷil*; VRS *tøβusiŋmʷil*; MTA *taβusiŋmʷele*; LKN *taβusiŋmʷel*

The word’s etymology is tricky. In principle, it parses as \**ta-βusi-mʷele*, literally ‘one who has struck a Cycas palm’. However, the element \**mʷele* – while originally the name of the highly prestigious palm – does not designate here the plant itself, but stands for the high rank of the same name (87):

- (87) \**mʷele* ‘Cycas palm’ → ‘name of a chiefly rank, relatively high’: HIW *ŋʷij*; LTG *ŋʷil*; VLW *ɛy̩le|ŋ̩mʷel*; MTP *ne-ŋ̩mʷel*; MTA *ŋ̩mʷele*

The underlying object of \**βusi* ‘strike, hit’ is not \**mʷele*, but an implicit ‘pig’, whose sacrifice is necessary to attain higher ranks. Thus the etymology of \**taβusimwele* is really ‘a man who has killed [pigs] for the Cycas chiefly rank’ (Codrington 1891: 55, 1896: 211) – hence ‘a man of great rank’.

While the chiefly positions were restricted to men, some women also were regarded highly – either because they were the wives of high chiefs, or because they had gone through their own hierarchy of grades (Hess 2009: 10). The term then used is a reflex of \**mʷotari*:

- (88) \**mʷotari* ‘woman of high prestige’  
 → HIW *ŋʷətɔ̩gL*; MTP *jɔ-ŋ̩mʷɔtej*; VRS *ŋ̩mʷeter*; MTA *ŋ̩mʷotar*; LKN *ŋ̩mʷɔtæ:*; MRL *ne-ŋʷeter*

The sons and nephews of a highly-ranked chief would also have prestige. It is possible to reconstruct a protoform \**woyutu* with this meaning:

- (89) \**woyutu* ‘young man of high prestige; admirable man’  
 → HIW *wowut*; MTP *nu-wɔyit*; VRS *wøwut*; MTA *wowut*; KRO *wuwut*; LKN *wuwut*

Some of my consultants would liken this word with the western notion of ‘Prince’. In Hiw, the two words *ŋʷətɔ̩gL* and *wowut* have now taken on a profane meaning, that of an admirable woman and admirable man, respectively.

## 4.8 Supernatural powers

### 4.8.1 \**mana* ‘supernatural power’

The social aura which surrounds high chiefs in the ancient political system, but also shamans and sorcerers, is inseparable from a notion of supernatural power, called \**mana*:

- (90) \**mana* → [N] ‘supernatural power held by a person or a thing; magic force’: HIW *manə*; LTG *menə*; LHI *n-man*; LYP *n-man*; VLW *n-man*; MTP *na-man*; LMG *n-mpn*; VRA *mana*; VRS *man*; MSN *man*; MTA *mana*; NUM *man*; LKN *man*; MRL *na-man*

This term *mana* (<POc \**manan*), widespread in the Pacific, has been widely studied by anthropologists working in that part of the world (Codrington 1891; Ivens 1931; Keesing 1984; Mondragón 2004). In various Oceanic languages, *mana* is a verb meaning ‘be efficacious, be true, be potent’ (Keesing 1985: 203); yet in northern Vanuatu, the use of *mana* as a verb is marginal: it is only found in a few languages (Mwotlap, Vurës, Mota), typically in phrases restricted to seafood, meaning ‘[fish, food] be poisonous’. In the Torres–Banks area, the principal use of \**mana* is as a noun, referring to a magical force that is present in certain places, objects, or individuals. A healer, sorcerer, magician either “has” or “lacks” \**mana*: e.g. in Mwotlap *na-man ai mi ki* (literally ‘there is some *mana* with him’, i.e. ‘he has *mana*’). A sacred stone, a haunted place, a valued name may also bear their own *mana*.

In itself, the power of \**mana* is amoral, and can be put to benevolent or malevolent use. Thus, a good poet is supposed to have \**mana*, a force that gives him his talent and inspiration; this \**mana* is transmitted from master to disciple, during a special ritual involving magic leaves. A high chief in the \**sumbwe* would also have been endowed with considerable \**mana*, as he found the power to pursue his political career (Vienne 1984: 377); the term is not used when referring to the modern, secular forms of chiefdom.

But the word *\*mana* is also often associated with the awe and fear evoked by sorcery. Most languages use *\*mana* to refer to the dangerous, supernatural power of ancestral and evil spirits. In Lo-Toga, *menə hia* (lit. ‘evil mana’) designates the practice of sorcery itself. Mwotlap has a derived verb *manhey* [sorcerer] cast a spell on s.o.’.

#### 4.8.2 \**lalaβi* ‘harmful aura of evil spirits’

The languages of the area have other words to refer to the more or less magic power associated with a person or thing. One of them is *\*lalaβi*:

- (91a) \**lalaβi* → ‘harmful aura emanating from an evil spirit’:

LMG *n-lvlβi*; DRG *lalβi*; KRO *lalβi*; LKN *lalβi*

I have mostly found this word in relation to evil spirits *\*[a]tamate*, in a phrase like LMG *n-lvlβi ?ama?* ‘the evil power of phantoms’ – a meaning which other languages would express with *\*mana*. The word *\*lalaβi* has an interesting etymology, which can be reconstructed thanks to the conservative language Mota (Codrington and Palmer 1896: 43):

- (91b) MTA *lalaβi-* → ‘the blast of heat; the blast of a swift passing body, as a falling rock’

This noun is itself derived from a verb *lalaβ* ‘send out a blast, burn, scorch’, in turn linked to *lala*:

- (91c) \**lala* → ‘[fire] burn clear and hot; smoulder’:

HIW *jaj*; LYP *lal*; VLW *lal*; MTP *lal*; LMG *lvl*; MTA *lala*

In sum, the word *\*lalaβi* likens the harmful power of evil spirits to the scorching heat of a burning fire.

#### 4.8.3 \**tayaro* ‘supernatural force embodied in spirit and stones’

Finally, another word sometimes used to refer to someone’s magical power is *\*tayaro* (forms are here given with a 3sg possessor: ‘his/her/its power’):

- (92a) \**tayaro* (-na) → ‘supernatural power emanating from a person, a spirit or a force’:

LMG *n-?ayro-n ~ n-?ayar*; VRS *tayro-n*

The semantics of this word family is puzzling, and deserves to be explored. Thus in Lemerig, the term *?ayar* was defined as a supernatural power found with non-human spirits, or with “superior men” such as initiate men and high chiefs. By contrast, a Vurës speaker, faced with this Lemerig definition, explained that the term *tayro-n* in Vurës could only refer to the magic force of a person (sorcerer, healer), but not of a spirit.

Other reflexes of *\*tayaro* do not even have any connection with the notion of magic power. For example, Vurës has an etymological doublet *tayar* ‘(good or bad) outcome of s.o.’s behaviour, (good or bad) luck’, which is restricted to human, non-magic referents. Likewise, Mwotlap has a reflex *na-tyaj* ‘fatigue, exhaustion’, perhaps via a meaning ‘effect of s.o.’s actions’.

So, if there was a semantic shift, was it from the sacred to the profane, or the reverse? In fact, it looks like this root *\*tayaro* is primarily linked to the realm of magic and supernatural forces; the profane senses found in Vurës and Mwotlap would result from more recent semantic shift. Indeed, besides the senses cited in (92a), the same root shows various meanings which are all related to supernatural powers. Thus, along with a reflex of *\*ndule* (59b), Lo-Toga has a second term for a slightly different sort of haunted stone:

- (92b) \**tayaro* → ‘magic stone in the bush, used as altar for sorcery’: LTG *tøyar*

And the neighbouring Hiw has a related verb *tegLyje* ‘bewitch’, used in black magic. This form is the regular reflex of a protoform *\*tayaro-a*, i.e. a derived verb bearing the transitivising suffix *\*-a* (cf. 76):

- (92c) \**tayaro-a* → ‘bewitch, cast a spell on s.o. using sorcery’: HIW *tegLyje*

All these elements tend to confirm that the root *\*tayaro* was primarily associated with mythical figures and supernatural forces. Given the variety of its modern reflexes, it is difficult to assign any precise meaning to the etymon. The common denominator between all these senses could perhaps be defined as *\*tayaro* ‘a supernatural force embodied in a non-human spirit or a magic stone, and invoked by a sorcerer in bewitching his victims’. In line with Figure 4 above, Figure 5 charts the various senses which northern Vanuatu languages associate with this root.

Finally, Mota (Codrington and Palmer 1896) also has three words reflecting the same root (albeit with an irregular sound correspondence *\*y-\**η):

- *tanaro* ‘a morsel of food thrown for a ghost’
- *tanaro-na* ‘a thing belonging to a person which has magic power’
- *tanaroa* ‘stone, carried or hung up in a bag, possessed of magic powers as the abode of a *Vui* spirit’

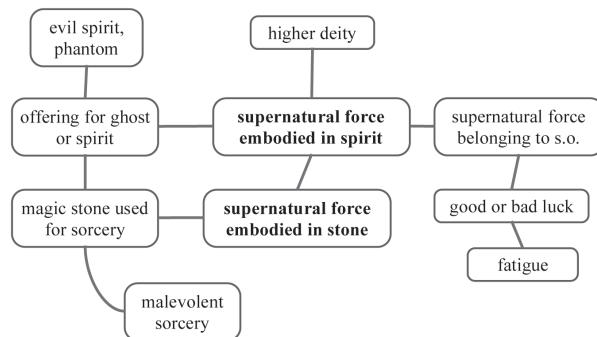


Figure 5: A lexical map for *\*tayaro* 'supernatural force'

The latter form *tajaroa* is clearly the same word as Dorig/Lakon *tajro* 'evil spirit, phantom, ogre' mentioned in §4.4.2. Crucially, it is also homophonous with *Tajaroa*, the name of the famous Polynesian deity. In fact, we saw that northern Vanuatu languages have a deity *\*Tayaro* (57), the brother of the main deity Qat. Jolly (1996: 241) also mentions *tegar* as the name of an evil spirit in southern Pentecost, and explicitly points out (p. 257) the similarity with the names of the divinities Tagaro and Tajaroa. In my view, there is no strong reason to hypothesize any Polynesian borrowing here. Given the amount of sound change that has affected its form, and also the deep-reaching connections with other supernatural meanings (Figure 5), it is much more likely that the spiritual force *\*Tayaro* ~ *\*Tajaro* was already present in the mythology of the first Oceanic settlers who peopled Vanuatu. If this is true, then *\*Tajaro* (the form reconstructable at the level of POC) would potentially be the first example of an ancient god whose existence could be reconstructed for the early times of Proto Oceanic unity.

## 5 Conclusion

Due to three millennia of sound change and semantic shift, many words of northern Vanuatu languages have altered their shape or their meaning over time, and have often lost any obvious link with each other. However, should one pay careful attention to sound correspondences so as to avoid false hypotheses, comparison across modern languages makes it possible to retrieve these historical connections between otherwise separate concepts.

This is how etymological research, applied to the languages of northern Vanuatu, has proven able to unearth the historical links that connect 'mind' and 'inside' (\**lolo*); 'forget' and 'night' (\**m̥b̥wɔŋi*); 'soul' and 'shadow' (\**ata*); 'ghost' and 'secret society' (\**atamate*); 'political system' and 'god' (\**su<sup>m</sup>b̥wē*); 'deity' and 'root' (\**βui*); 'taboo' and 'cemetery' (\**ta<sup>m</sup>bu*); 'Underworld' and 'volcano' (\**m̥b̥anoi*) or 'cave' (\**asura*); 'world' and 'light' (\**marama*) ... – to cite but a few. All these connections have now become opaque to the speakers themselves, and would have been difficult to detect solely based on language-internal analyses: they can only be discovered through a methodical process of cross-linguistic comparison.

Like the spirits who surround the living, the words we speak today are the shadows of bygone lives. The science of etymology provides the linguist with the shamanic power to see the invisible, and bring back to the light of our present world the vanished souls of our ancestors.

## References

- Benveniste, Émile. 1969. *Le Vocabulaire des Institutions Indo-européennes*. 2 volumes. Le Sens Commun. Paris: Minuit.
- Biggs, Bruce. 2006. *Kimihia te mea ngaro: Seek that which is lost* (Memoirs of the Polynesian Society 53). Auckland: Polynesian Society.
- Blust, Robert. 1972. Note on PAN \*qa(R)(CtT)a 'outsiders, alien people'. *Oceanic Linguistics* 11. 166–171.
- Blust, Robert. 1980. Austronesian Etymologies. *Oceanic Linguistics* 19 (1/2). 1–189.
- Blust, Robert. 1981. Linguistic Evidence for Some Early Austronesian Taboos. *American Anthropologist* 83 (2). 285–319.
- Blust, Robert. 1987. Lexical reconstruction and semantic reconstruction: the case of Austronesian 'house' words. *Diachronica* 4. 79–106.
- Blust, Robert. 2009. *The Austronesian languages* (Pacific Linguistics 602). Canberra: Australian National University.
- Blust, Robert. 2012. *Austronesian Comparative Dictionary*, web edition. <http://www.trussel2.com/ACD> (accessed 1 June 2012).
- Bolton, Lissant. 2003. *Unfolding the moon: Enacting women's kastom in Vanuatu*. Honolulu: University of Hawaii Press.
- Charnraine, Pierre. 1968–1980. *Dictionnaire étymologique de la langue grecque: Histoire des mots*. 4 volumes. Paris: Klincksieck.
- Charpentier, Jean-Michel & Alexandre François. forthcoming. *Atlas Linguistique de Polynésie Française – Linguistic Atlas of French Polynesia*. 2 volumes. Berlin: Mouton de Gruyter.
- Clark, Ross. 2009. *\*Leo Tuai: A comparative lexical study of North and Central Vanuatu languages* (Pacific Linguistics 603). Canberra: Australian National University.
- Codrington, Robert H. 1885. *The Melanesian languages*. Oxford: Clarendon Press.

- Codrington, Robert H. 1891. *The Melanesians: Studies in their anthropology and folklore*. Oxford: Clarendon Press.
- Codrington, Robert H. & John Palmer. 1896. *A dictionary of the language of Mota, Sugarloaf Island, Banks Islands*. London: Society for Promoting Christian Knowledge.
- Crowley, Terry. 2000. The language situation in Vanuatu. *Current Issues in Language Planning* 1 (1). 47–132.
- Dahl, Otto Christian. 1981. *Early phonetic and phonemic changes in Austronesian* (Institute for Comparative Research in Human Culture series B-63). Oslo: Norwegian University Press.
- Dempwolff, Otto. 1969 [1938]. *Vergleichende Lautlehre des Austronesischen Wortschatzes: Austronesisches Wörterverzeichnis*, 2nd edn. Liechtenstein: Nendeln.
- Fox, Charles Elliot. 1962. *Kakamora*. London: Hodder and Staughton.
- Fox, Charles Elliot & F. H. Drew. 1915. Beliefs and Tales of San Cristoval. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland* 45. 187–228.
- François, Alexandre. 2000a. Vowel shifting and cloning in Motlav: historical explanation vs. formal description. In: M. Klamer (ed.), *Proceedings of AFLA7, the 7th Meeting of Austronesian Formal Linguistics Association*, 49–68. Amsterdam: Vrije Universiteit.
- François, Alexandre. 2000b. Déivation lexicale et variations d'actance: petits arrangements avec la syntaxe. *Bulletin de la Société de Linguistique* 95 (1). 15–42.
- François, Alexandre. 2002. *Araki. A disappearing language of Vanuatu* (Pacific Linguistics 522). Canberra: Australian National University.
- François, Alexandre. 2005. Unraveling the history of the vowels of seventeen northern Vanuatu languages. *Oceanic Linguistics* 44 (2). 443–504.
- François, Alexandre. 2007. Noun articles in Torres and Banks languages: Conservation and innovation. In: John Lynch, Jeff Siegel and Diana Eades (eds.), *Language Description, History and Development: Linguistic Indulgence in Memory of Terry Crowley* (Creole Language Library 30), 313–326. Amsterdam & Philadelphia: John Benjamins.
- François, Alexandre. 2008a. *Rythmes à danser, poèmes à chanter en Mélanésie: Esthétique, transmission et impact social des arts musicaux au Vanuatu*. Final report for the grant “Jeunes Chercheurs” (2004–2007). Paris: LACITO, Centre National de la Recherche Scientifique.
- François, Alexandre. 2008b. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In: Vanhove, Martine (ed.), *From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic Associations* (Studies in Language Companion Series 106), 163–215. Amsterdam & Philadelphia: John Benjamins.
- François, Alexandre. 2009. The languages of Vanikoro: Three lexicons and one grammar. In: Bethwyn Evans (ed.), *Discovering history through language: Papers in honour of Malcolm Ross* (Pacific Linguistics 605), 103–126. Canberra: Australian National University.
- François, Alexandre. 2010. A comparison of geocentric directionals across Banks and Torres languages. Paper read at the 8th International Congress of Oceanic Linguistics (cool8), University of Auckland (NZ), 4–9 January.
- François, Alexandre. 2011a. Where \*R they all? The geography and history of \*R loss in Southern Oceanic languages. *Oceanic Linguistics* 50 (1). 140–197.
- François, Alexandre. 2011b. Social ecology and language history in the northern Vanuatu linkage: A tale of divergence and convergence. *Journal of Historical Linguistics* 1 (2). 175–246.
- François, Alexandre. forthcoming. *Voices from the Volcano: Myths, Folktales and Poetry from Gaua, Vanuatu* (World Oral Literature Series). Cambridge: Open Book Publishers.
- François, Alexandre. in prep. Systems of geocentric space reference in the languages of northern Vanuatu. In: Alexandre François, Sébastien Lacrampe, Stefan Schnell & Michael Franjeh (eds.), *The Languages of Vanuatu: Unity and Diversity* (Studies in the Languages of Island Melanesia). Canberra: Asia Pacific Linguistics Open Access.
- François, Alexandre & Monika Stern. forthcoming. *Vanuatu: Celebrations and Mysteries* (Inédit). Paris: Maison des Cultures du Monde.
- Hess, Sabine. 2009. *Person and Place: Ideas, Ideals and the Practice of Sociality on Vanua Lava, Vanuatu*. Oxford: Berghahn Books.
- Hyslop, Catriona. n.d.a *Vera'a–English lexicon*. Melbourne: Research Centre for Linguistic Typology. Unpublished lexical files.
- Hyslop, Catriona. n.d.b *Vurës–English lexicon*. Melbourne: Research Centre for Linguistic Typology. Unpublished lexical files.
- Ivens, Walter G. 1931. The Place of Vui and Tamate in the Religion of Mota. *Royal Anthropological Institute of Great Britain and Ireland* 61. 157–166.
- Jolly, Margaret. 1996. Devils, Holy Spirits, and the Swollen God: Translation, Conversion and Colonial Power in the Marist Mission, Vanuatu, 1887–1934. In: Peter van der Veer (ed.), *Conversion to Modernities: the Globalization of Christianity*, 231–262. New York, London: Routledge.
- Keesing, Roger. 1984. Rethinking *mana*. *Journal of Anthropological Research* 40. 137–156.
- Keesing, Roger. 1985. Conventional metaphors and anthropological metaphysics: The problematic of cultural translation. *Journal of Anthropological Research* 41 (2). 201–217.
- Koptjevskaja-Tamm, Maria, Martine Vanhove & Peter Koch. 2007. Typological approaches to lexical semantics. *Linguistic Typology* 11 (1). 159–186.
- Lanouguère-Bruneau. 2002. *Le corps de l'igname et le sang de la noix de coco: le système social ancien de Mwotlap (îles Banks, Vanuatu)*. Paris: École des Hautes Études en Sciences Sociales dissertation.
- Lynch, John. 2002. Potent Roots and the Origin of kava. *Oceanic Linguistics* 41 (2). 493–513.
- Lynch, John & Terry Crowley. 2001. *Languages of Vanuatu: A new survey and bibliography* (Pacific Linguistics 517). Canberra: Australian National University.
- Marck, Jeff. 1996. The first-order anthropomorphic gods of Polynesia. *Journal of the Polynesian Society* 105 (2). 217–258.
- Mondragón, Carlos. 2004. Of Winds, Worms and Mana: The Traditional Calendar of the Torres Islands, Vanuatu. *Oceania* 74 (4). 289–308.
- Pawley, Andrew. 2003. Locating Proto Oceanic. In: Malcolm Ros; Andrew Pawley & Meredith Osmond (eds.). 1998. *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society. Vol. 1: Material culture* (Pacific Linguistics C-152), 17–34. Canberra: Australian National University.
- Pawley, Andrew. 2005. The meaning(s) of Proto Oceanic \*panua. In: Claudia Gross, Harriet D. Lyons & Dorothy A. Counts (eds.), *A Polymath Anthropologist: Essays In Honour of Ann Chowning*. Research in Anthropology and Linguistics, Monograph 6. University of Auckland: Department of Anthropology.
- Pawley, Andrew. 2007. The origins of Early Lapita culture: the testimony of historical linguistics. In: Stuart Bedford, Christophe Sand & Sean P. Connaughton (eds.), *Oceanic Explorations: Lapita and Western Pacific Settlement* (Terra Australis 26), 17–50. Canberra: ANU E-Press.

- Pawley, Andrew. 2010. Prehistoric Migration and Colonisation Processes in Oceania: A View from Historical Linguistics and Archaeology. In: Jan Lucassen (ed.), *Migration history in world history: multidisciplinary approaches* (Studies in Global Social History 3), 77–112. Leiden: Brill.
- Pawley, Andrew & Malcolm Ross, eds. 1994. *Austronesian Terminologies: Continuity and Change*. Pacific Linguistics C-127. Canberra: Australian National University.
- Ross, Malcolm. 1988. *Proto-Oceanic and the Austronesian languages of Western Melanesia*. Pacific Linguistics. Canberra: Australian National University.
- Ross, Malcolm, Andrew Pawley & Meredith Osmond (eds.). 1998. *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society. Vol. 1: Material culture* (Pacific Linguistics C-152). Canberra: Australian National University.
- Ross, Malcolm, Andrew Pawley & Meredith Osmond (eds.). 2003. *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society. Vol. 2: The physical environment* (Pacific Linguistics 545). Canberra: Australian National University.
- Ross, Malcolm, Andrew Pawley & Meredith Osmond (eds.). 2008. *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society. Vol. 3: Plants* (Pacific Linguistics 599). Canberra: Australian National University.
- Ross, Malcolm, Andrew Pawley & Meredith Osmond (eds.). 2011. *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society. Vol. 4: Animals* (Pacific Linguistics 621). Canberra: Australian National University.
- Ross, Malcolm, Andrew Pawley & Meredith Osmond (eds.). in prep a. *The lexicon of Proto Oceanic. People: Body and mind* (Pacific Linguistics). Canberra: Australian National University.
- Ross, Malcolm, Andrew Pawley & Meredith Osmond (eds.). in prep b. *The lexicon of Proto Oceanic. People: Society* (Pacific Linguistics). Canberra: Australian National University.
- Traugott, Elizabeth C. & Richard B. Dasher. 2002. *Regularity in semantic change* (Studies in Linguistics 96). Cambridge: Cambridge University Press.
- Tryon, Darrell. 1976. *New Hebrides Languages: An internal classification* (Pacific Linguistics C-50). Canberra: Australian National University.
- Vanhove, Martine (ed.). 2008. *From Polysemy to Semantic change: Towards a Typology of Lexical Semantic Associations* (Studies in Language Companion Series 106). Amsterdam & Philadelphia: Benjamins.
- Vienne, Bernard. 1972. Les formes du pouvoir et de l'autorité dans les communautés paysannes des îles Banks (Nouvelles-Hébrides). Nouméa, Nouvelle-Calédonie: ORSTOM. Unpublished document.
- Vienne, Bernard. 1984. *Gens de Motlav. Idéologie et pratique sociale en Mélanésie* (Publication de la Société des Océanistes 42). Paris: Société des Océanistes.
- Vienne, Bernard. 1996. Masked faces from the country of the Dead. In: Joël Bonnemaison, Kirk Huffman, Christian Kaufmann & Darrell Tryon (eds.), *Arts of Vanuatu*, 234–246. Bathurst: Crawford House Press.



# ***Temperature terms in Vanuatu languages***

by **Alexandre FRANÇOIS**

LACITO-CNRS

Australian National University

---

*Oceanic languages generally have a small set of lexemes encoding temperature predicates - no more than two or three words in each language. Focusing on a sample of languages from northern Vanuatu, this study describes their temperature terms, analyses the syntax of their various case frames, and highlights relevant semantic contrasts. Their temperature lexicons in this area appear to display a limited array of polysemies across semantic domains - whether compared to other languages of the world, or to other semantic domains in the same languages.*

---

\*

One of the purposes of this collective volume is to describe the cross-linguistic diversity in the domain of temperature terms, and their various semantic extensions. A fair part of this diversity takes the form of elaborate polysemies and metaphors that can be observed in some parts of the world while not in others. The present short study tackles diversity from a different angle - by reporting on a region, Vanuatu, where temperature terms show few significant semantic extensions. While this report promises to provide little substantial contribution to the typology of temperature polysemies, its interest lies precisely in the “negative evidence” it brings to the debate: should anyone claim the universality of certain semantic extensions, then this empirical study should illustrate a case where no such extensions are found.

I will first situate the languages of northern Vanuatu in their geographical and social context, including notes on the experience of temperature in this part of the world (Section 2). After presenting the nature of my data, a brief section will mention some key aspects of grammar, relevant to this study (Section 3). This will be followed by observations on major temperature terms and the various case frames in which they are used (Section 4). I will show, for example, that the experiencer of temperatures is treated sometimes as the syntactic subject, sometimes as the object, and is occasionally left out altogether.

The remainder of the article will delve into the semantic properties of the temperature lexicon, including their etymologies, their semantic contrasts, and their polysemies when they exist (Section 5). I will argue that the Oceanic languages here studied show relatively little by way of semantic extension – both by contrast with other parts of the world, and by language-internal comparison with other semantic domains.

## 1 The languages of Vanuatu

One of the archipelagoes of Island Melanesia, the modern country of Vanuatu – known as the New Hebrides until its independence in 1980 – hosts the world's highest density of languages *per capita* (Crowley 2000). A total of 106 vernacular languages have been recorded there (Tryon 1976; Lynch & Crowley 2001) for a current population of 234,000 inhabitants. The population lives scattered in the country's various islands, where for the most part they pursue traditional activities of farming and fishing, in line with the way of life of their ancestors during the last 3000 years of settlement in the region. *Figure 1* situates Vanuatu within Island Melanesia and the Pacific. The present study will focus on the northernmost area of Vanuatu, the island groups known as the Torres Islands and the Banks Islands.

*Figure 2* provides a linguistic map of the Torres and Banks Islands.

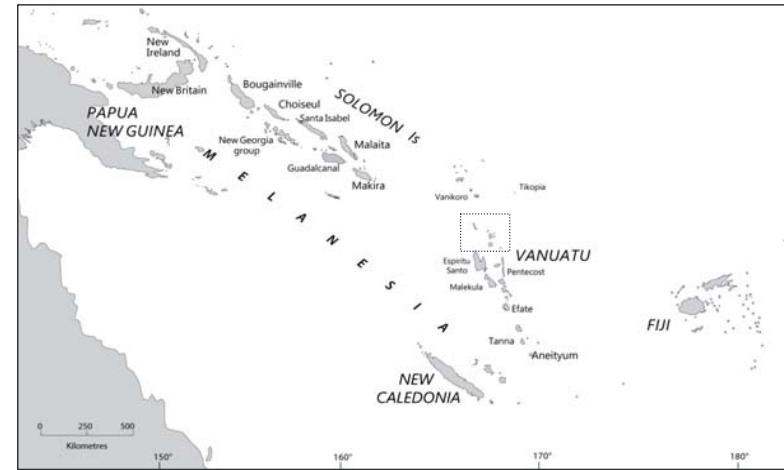


Figure 1 – Situation of Vanuatu, and in particular Torres-Banks islands, in the Pacific

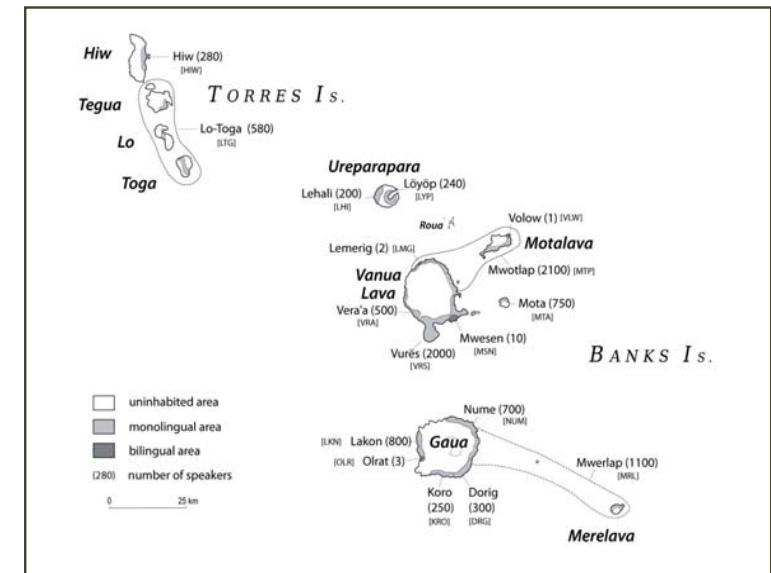


Figure 2 – The 17 languages of northern Vanuatu (Torres and Banks islands)

All these languages share a common ancestor, Proto Oceanic – itself a member of the vast Austronesian family. In a way similar to the fragmentation of Latin into a multitude of Romance languages and dialects, the linguistic diversity observed today in Vanuatu results from three millennia of *in situ* diversification from what was once a single language spoken across a vast social network (François 2011a, 2011b).

Even though the languages of northern Vanuatu have long diverged so as to lose mutual intelligibility, their linguistic structures display considerable similarities, due to their shared ancestry as well as to sustained relations of contact (François 2011b). This is true of their grammars – which show massive areal isomorphism – but also of their lexicons, whose structural organisation is usually parallel across neighbouring languages. Even when historical processes of lexical replacement have resulted in different word *forms* across neighbouring languages, these words' semantic outlines and polysemies typically align from one language to the other. This point will be relevant to this survey of temperature terms: it will mean that most statements that can be made about the lexical semantics of a sample of languages can generally be considered representative of the whole area. Of course, this general principle of widespread lexical isomorphism does not preclude the possibility that certain lexical structures can be specific to just some languages – as will appear below.

The present study rests on primary data collected by the author, during a number of field trips in northern Vanuatu since 1997. These surveys involved the study of 18 different languages – i.e. the 17 languages of the Torres and Banks Islands, plus Araki further south.<sup>1</sup> Publications on these languages include a grammar and lexicon of Araki (François 2002, 2008b), a description of Mwotlap (François

<sup>1</sup> For each language, I collected lexical and grammatical information, and recorded a number of stories. My total Vanuatu corpus includes 660 text recordings in these 18 languages, with a total of 73 hours. Among these, the best documented language is Mwotlap, with 314 recordings among which 110 were transcribed, yielding a text corpus of over 100,000 words.

2001, 2005), as well as a number of works currently in preparation: a dictionary of Mwotlap; a lexicon of Lo-Toga; a grammar of Hiw; text collections in the languages of Gaua.

## 2 Experiencing temperature in Vanuatu

The archipelago of Vanuatu is located close to the equator, ranging in latitude from 13° to 20° south. Its sub-tropical climate is characterised by approximately eight months – from October to May – of warm to hot rainy weather with frequent cyclones. Around the month of June, that wet season is followed by a relatively drier season of about four months when the weather cools down, thanks to the trade winds blowing from southeast. Depending on the season and on the time of the day, the breeze blowing from the sea can be felt to be warm, or refreshing.

Daily temperatures are relatively stable during the year, ranging on average from +20 °c to +32 °c. The temperature of the sea ranges from +22 °c in winter to +28 °c in summer. The experience of weather temperatures is essentially limited to this rather narrow range. One context where people can feel cold, though, is at high altitude: this is especially the case on volcanic islands (e.g. Vanua Lava or Gaua in the north, Santo or Tanna further south) where some villages are located inland, up to about 900 m in altitude. Temperatures there can go down dramatically to about +13 °c at night, setting the local standard for extreme cold.

Another context where relative cold can be experienced is simply when, during a very hot moment, some thermal contrast can be found in the shade, a sudden rain, or by bathing in cool water. Houses are traditionally made with walls of bamboo and with a thatched roof made of sago leaves; these materials are efficient at keeping temperatures cool even when outside temperatures are high. Natural caves tend to have much cooler temperatures, yet they are only inhabited when people take shelter during hurricanes. As we shall see, local lexicons have no separate word for 'cool': because time in

the shade or contact with water are the only moments when bodies can feel a drop in temperatures, these will always be described simply using the word for ‘cold’ – generally with a positive connotation.

Traditional cultures have no technology for cooling down anything below ambient temperatures. Fridges or ice are unknown in the rural areas of the country, where electricity is rare and intermittent. Technology for increasing temperatures revolves around the use of fire. Except for some occasional cool moments in the mountains, fire is hardly ever used as a way for people to warm up; its main function is for providing light, cooking food, or boiling water. While traditional ways of cooking also include roasting or frying, the main techniques involve the “earth oven”. This is an 80-cm-wide round pit dug in the ground, in which a fire is lit so as to heat a number of stones; once red-hot, these stones will store heat for several hours, and form a hot bed above which food will be laid, covered with leaves, and baked. This way of cooking, common throughout the Pacific (and known in Polynesia as *umu*) sometimes involves the production of hot steam, as water – often salt water from the sea – is poured directly onto the hot stones so as to steam food rather than bake it.

Temperature as such is not involved prominently in any social event or ritual that I know of. The main potential connection between temperature and social activities would be through fire and cooking. Indeed, social events such as weddings, village celebrations, inaugurations or farewell parties, invariably involve a collective meal as a key element in the unfolding of events. As a corollary, major social events often involve gatherings the night before, so as to prepare large quantities of food for the next day. Food would be prepared, and the earth oven lit up, so as to bake the whole night long. Entire families would stay up all night in the vicinity of the oven – not too close to avoid the smoke or the steam – and tell stories or spend time together. It is frequent to hear phrases that associate metonymically village celebrations with the domain of fire and cooking. Thus on Motalava island, as a young man addresses his prospective parents-

in-law with respect and modesty, the elliptic phrase “*I'd like to go fetch some firewood*” sometimes suffices to evoke the whole wedding ceremony, and thus functions effectively, in a Gricean kind of way, as a marriage proposal.

### 3 Preliminary notes on grammar

This section presents some grammatical features that are shared among the languages of northern Vanuatu, and which are relevant to the discussion of temperature terms. Unless otherwise specified, most examples will come from the Mwotlap language (MTP).

#### 3.1 Parts of speech in northern Vanuatu languages

Northern Vanuatu languages all have accusative alignment, and a strict SVO word order (that is, SV and AVO). Core arguments are assigned case on the basis of word order only.<sup>2</sup>

(MTP.1)	no	n-ixlal	ki,	<sup>m</sup> ba	ki	εt=	ixlal	=tc	no.
	1sg	STAT-know	3sg	but	3sg	NEG <sub>1</sub> =	know	= NEG <sub>2</sub>	1sg

‘I know her, but she doesn't know me.’

The system of major parts of speech in these languages distinguishes between nouns, adjectives, verbs.

Unlike other parts of speech, nouns can be the head of a REFERENTIAL PHRASE. For that purpose, most nouns must be prefixed by an article *nV*- encoding referentiality: thus *lomxep* ‘boy’ becomes *no-lomxep* ‘a/the boy’ in (2).

(MTP.2)	ki	ni-etsas <sub>v</sub>	no-lomxep <sub>N</sub>	su <sub>ADJ</sub>	vitwax.
	3sg	AOR-see	ART-boy	small	one

‘He saw a small boy.’

Unlike verbs, adjectives can fill the function of modifier in a noun phrase – e.g. *su* ‘small, young’ in (2). As for verbs, they can neither

<sup>2</sup> Throughout this study, forms will be spelled phonemically using IPA conventions rather than local orthographies, to facilitate cross-linguistic comparison.

head a referential phrase, nor act as a direct noun modifier; their main function is to head a PREDICATE PHRASE, in which case they are inflected for Tense-Aspect-Mood – like *iygal* ‘know’ in (1), or *etsas* ‘see’ in (2).

The function of TAM-inflected predicate is open not only to verbs, but also to adjectives and nouns:

- (MTP.3)    ki    ni-in    ε    ki    ni-*su<sub>ADJ</sub>*    lok.  
           3sg    AOR-drink    COORD    3sg    AOR-small    again  
           'He drank it and became young again.'

- (MTP.4)    ki    ni-in    ε    ki    ni-*lomg̊ep<sub>N</sub>*    lok.  
           3sg    AOR-drink    COORD    3sg    AOR-boy    again  
           'He drank it and became a boy again.'

There are no reasons to consider that the adjective *su* ‘small’ in (3), or the noun *lomg̊ep* ‘boy’ in (4), have been converted into verbs. A more accurate formulation is to consider that these are two examples of TAM-predicates which are headed, not by verbs, but by (respectively) an adjective and a noun. Indeed, like most Austronesian languages, the languages of northern Vanuatu are “omnipredicative” (cf. Lemaréchal 1989:55, Launey 1994): that is, all major syntactic categories are equally compatible with the predicate function, with no need of any copula.

In sum, the syntactic function of *predicate* is open to nouns, adjectives and verbs alike, and cannot be taken as a criterion to tell them apart. The features which distinguish these syntactic categories have to be found outside the predicate, such as in the internal structure of referential phrases (NPs). Each syntactic category is characterised by its own set of syntactic compatibilities, which are summarised in *Table 1*. For a more detailed discussion of parts of speech in Mwotlap, see François (2001, 2003).

*Table 1* – Syntactic functions of major categories in Mwotlap

	<b>Noun</b>	<b>Adjct</b>	<b>Verb</b>
be head of predicate	+	+	+
modify head of predicate	+	+	+
be head of referential phrase	+	-	-
modify head of ref. phrase	+	+	-

### 3.2 The polyfunctionality of reduplication

One morphological feature shared by northern Vanuatu languages, and by Oceanic languages in general, is the common use of reduplication to encode a variety of meanings. Among its many uses (François 2004a), I will mention briefly the ones most relevant for the understanding of the sentences cited in the next sections – based on Mwotlap examples.

Certain nouns, adjectives or verbs optionally use reduplication to encode plurality:

- *lomg̊ep* [N] ‘boy’ → *lomlomg̊ep* [N] ‘boys’
- *su* [ADJ] ‘small:SG’ → *susu* [ADJ] ‘small:PL’
- *mat* [V] ‘die’ → *matmat* [V] ‘die in numbers’

Reduplication can also encode intensity, particularly with adjectives:

- *su* [ADJ] ‘small’ → *susu* [ADJ] ‘very small, tiny’
- *sew* [ADJ] ‘hot’ → *sewsew* [ADJ] ‘very hot’

We will see various examples of intensifying reduplication with *sewsew* in our next sections, e.g. (40)-(41). The form will then be glossed ‘hot~INTSF’, using the tilde sign recommended by the Leipzig glossing rules for reduplication.

Among its possible aspectual meanings, reduplication encodes progressive or imperfective aspect – as in ex. (19) below with *sewsew* (‘hot~IPFV’, i.e. ‘was heating up’).

Finally, reduplication is also used in causative serial constructions, when the second lexeme (verb, adjective) has resultative meaning. For example, *təŋ* means ‘cry’, but it is reduplicated when used as

a resultative predicate in a verb serialising pattern:

- (MTP.5)    imam    nənən    mu-wuh    *teŋteŋ*    kr.  
           father    his/her    PFT-slap    cry~RESULT    3sg  
           'Her father made her cry by slapping her.'  
           [lit. 'slapped her to cry']

We will see below examples of resultative reduplication with *səwsew*, this time glossed 'hot~RESULT' – e.g. (11) or (43).

Even though reduplication is highly productive for a large part of the lexicon, it is not available for all lexemes. Indeed, it is common for a given root to have historically lost its simplex form, and become lexicalised in its reduplicated form; most of the time, the resulting (formerly reduplicated) word has become immune to any further morphological reduplication. In fact, several of the temperature words cited in *Table 2* below are in this case. Etymologically, the adjective \**tutunu* 'hot' results from the reduplication of a verb \**tunu* 'heat up, roast' (§5.1). But the reflex of this protoform \**tutunu* in modern languages – for example *?i?in* in Lemerig – has become insensitive to reduplicative morphology: it lacks a simplex form \**i?in*, and also lacks the ability to reduplicate further (\**i?i?i?in*). This principle applies to most forms in *Table 2*, whether they reflect \**tutunu* 'hot' or \**mamadridri* 'cold', as both result from earlier processes of reduplication. As a consequence, while Mwotlap *səw* 'hot' – as we just saw – is highly sensitive to reduplicative morphology, its counterpart *məmjjij* 'cold' (a regular reflex of \**mamadridri*) is unable to reduplicate.

## 4 Form and syntax of major temperature terms

### 4.1 Major temperature terms

The most relevant terms for our discussion of temperature lexicons in northern Vanuatu are adjectives. *Table 2* provides the forms for 'cold' and 'hot' in the languages of the Torres and Banks Islands. While

several languages in the corpus have only these two temperature words, some also have an additional term (see §5.2 below).

*Table 2* – Temperature-related lexemes  
in Torres and Banks languages

	'hot'	'cold'	other words
<b>Hiw</b>	ttin	maetit	+ ADJ <i>məwmwɪw</i> 'cold'
<b>Lo-Toga</b>	tən	meləhīh	+ NOUN <i>məla</i> 'cold'
<b>Lehali</b>	ssəw	mam <sup>n</sup> din	
<b>Löyöp</b>	ssəw	mam <sup>n</sup> din	
<b>Mwotlap</b>	<i>səw</i>	<i>məmjjij</i>	
<b>Volow</b>	<i>səw</i>	<i>mamjjij</i>	
<b>Lemerig</b>	<i>?i?in</i>	<i>mamwit</i>	+ ADJ <i>vævæt</i> 'hot'
<b>Vera'a</b>	<i>?i?in</i>	<i>mamgjin</i>	
<b>Vurës</b>	<i>tütün</i>	<i>mamixin</i>	+ NOUN <i>malas</i> 'cold'
<b>Mota</b>	<i>tutun</i>	<i>mamarir</i>	+ NOUN <i>malaso</i> 'cold'
<b>Nume</b>	<i>ttun</i>	<i>mamirir</i>	+ NOUN <i>malas</i> 'cold'
<b>Dorig</b>	<i>ttun</i>	<i>mamrir</i>	
<b>Lakon</b>	<i>tutun</i>	<i>misij</i>	
<b>Mwerlap</b>	<i>səw</i>	<i>mmerir</i>	+ VB <i>malas</i> 'feel cold'

## 4.2 Syntax and phraseology

### 4.2.1 ATTRIBUTIVE VS PREDICATIVE FUNCTION

Temperature adjectives are found with various syntactic functions, in conformity with the combinatorial possibilities of adjectives in these languages (§3.1). Thus Mwotlap *məmjjij* 'cold' is used attributively in (6):

- (MTP.6)    ni<sup>-m</sup>bi    *məmjjij*  
           ART-water    cold  
           'cold water'

In (7), *səw* 'hot' is used predicatively, with a TAM marker and an adverbial modifier:

- (MTP.7) na-lo nɛ-sɛw mɛh!  
 ART-SUN STAT-hot too.much  
 'The sun is too hot!'

As suggested in *Table 1* (§3.1), a third syntactic function open to adjectives is that of secondary predicate, that is, modifier of a verb: this will be exemplified in (10)-(11) or (45) below.

#### 4.2.2 STATIVE VS DYNAMIC INTERPRETATION

When combined with the stative aspect prefix *nV-*,<sup>3</sup> temperature adjectives in predicate position translate as 'BE cold/hot', a stative property assigned to an object at a given point in time – as in (7). Other TAM markers trigger a dynamic reading of the adjective: this is enough to make it refer to a change of state rather than to a property (François 2003:49). The predicate then translates as 'GET cold/hot', or using English intransitive (unaccusative) verbs, 'cool down/warm up'.

For example, the TAM marker in (8) is the "apprehensive", a type of detrimental modality. The mere use of a TAM other than the stative results in a change-of-state interpretation:

- (MTP.8) nr-<sup>m</sup>bi tiple mɔmɔjij lɔk.  
 ART-water APPREH cold again  
 'The water might cool down again.'

In (9), the predicate *sew* combines with the "aorist", whose various meanings include the marking of consecutive clauses:

- (MTP.9) nk̩ wr<sup>nd</sup>i hay nɛ-vɛt ki ni-sɛw εvej.  
 2sg (heap.up) up ART-stone 3sg AOR-hot INTSF  
 'Then you lay out the stones [on the earth oven]  
 until they become red-hot.'

As mentioned before, the syntactic functions open to adjectives include the modification of a predicate head – typically a verb. The adjective then forms with that verb a serial construction with a resultative meaning (François 2006:235), parallel to (5) above. This is

also a context that favours a dynamic (change-of-state) reading. Thus in (10), the complex predicate 'sit cold' means '[s.th.] remain in place until cooling down'. The construction is syntactically monovalent:

- (MTP.10) nɛ-ɣɛnɛŋen sɔ ni-hay mɔmɔjij tusu.  
 ART-food PROSP AOR-sit cold a.bit  
 'Let the food (rest and) cool down a little.'

In (11), the action is agentive, and the structure is that of a reflexive construction: that is, a formally transitive macro-verb in which the subject and the object are coreferent (François 2004b:118).

- (MTP.11) nɔk sɔ ñɔm<sup>w</sup>ihil sɛwsew nɔ van l-ɛp.  
 1sg PROSP dry.off hot~RESULT 1sg thither loc-fire  
 'I want to warm myself up (by standing) close to the fire.'  
 [lit. 'I will dry myself warm at the fire']

Notice here the use of reduplication *sɛwsew* to encode a resultative meaning (§3.2).

The languages under study lack any causative lexeme whose meaning would be directly linked to temperatures, like Eng. *heat (s.th.) up, cool (s.th.) down*. In the absence of any causative morphology, these meanings will be expressed by constructions such as (10)-(11), in which the adjectives *cold* and *hot* are used in resultative secondary predicates after a verb. We will encounter figurative uses of such causative constructions in §5.4.4.

#### 4.2.3 VALENCY AND CASE FRAME

The syntactic subject of temperature predicates is normally the carrier of temperature: that is, the theme to which this property is inherently assigned, whether a liquid (6), a solid (10) or a source of heat like fire or even the sun (7).

Northern Vanuatu languages normally do not allow *experiencers* to be the subject of a temperature predicate:

- (MTP.12) ?? nɔ nɔ-mɔmɔjij.  
 1sg STAT-cold  
 'I am cold.'

<sup>3</sup> The apparent homophony between the stative prefix *nV-* and the noun article *nV-* is a coincidence. It can be shown that the stative is underlyingly /ne-/ and the noun article /na-/ (François 2000:55).

I found only one exception to the latter principle: the verb *malas* in the language Mwerlap. This verb, which can be glossed '(s.o.) be cold, feel cold', is the only temperature-related predicate that can take the experiencer as its syntactic subject:

- (MRL.13) litneāk nu-mm̄erir lakan, ḡean nu-malas n̄eā.  
place STAT-cold DEIX 1inc:pl STAT-feel.cold OBL:ADV  
'It's cold here, that's why we *feel cold*.'

The Mwerlap verb *malas* is unique of its kind in the northern area. Even though *Table 2* includes cognate forms *malas* or *malaso*, in other languages this is a noun (§5.2.1), whose valency properties differ from the Mwerlap form.

Setting aside the Mwerlap exception, northern Vanuatu languages resort to other syntactic strategies to express the functional equivalent of a PERSONAL-FEELING construction (Kotpjievskaja-Tamm, 2011) such as Eng. *I am cold*. One strategy is to simply use an AMBIENT construction, which leaves the identity of the experiencer implicit. Weather-like statements referring to ambient temperature involve a dummy subject, which is always the noun for 'place, location'. An example was the first clause of (13) above for Mwerlap; see also (14) for Mwotlap, (15) for Hiw:

- (MTP.14) mahi n̄o-m̄omjij mesh!  
place STAT-cold too.much  
'It's too cold (here)!' [lit. the place is too cold]
- (HIW.15) n̄e ŋ̄utə ne ttin tñix!  
ART place STAT hot INTSF  
'It's very hot (here)!' [lit. the place is very hot]

This construction is parallel to statements about times of the day, for example:

- (HIW.16) n̄e ŋ̄utə ne k̄ej.  
ART place STAT night  
'It's night.' [lit. the place is night]

There are three ways to make the experiencer explicit in such statements. One is to embed ambient statements like (14) as a dependent clause under a verb of feeling:

- (MTP.17) n̄ok j̄ønt̄ej (mahi n̄o-m̄omjij).  
1sg feel place STAT-cold  
'I'm feeling cold.' [lit. 'I feel (the place is cold)']
- (HIW.18) sisə ɔ̄loŋ 〈n̄e ŋ̄utə ne ttin (mi sisə)〉.  
3pl feel ART place STAT hot with 3pl  
'They're feeling hot.'  
[lit. 'they feel (the place is hot (on them))']

When the language has a temperature-related noun, it can sometimes replace the object clause in a construction like (17): see the Vurës example (25) below.

A second construction is attested with the Mwotlap adjective *sew* 'hot'. Like a handful of other adjectives in this language, it can be used transitively. The subject is the source of the heat, while the object is the thing or person affected by the heat:

- (MTP.19) n̄-ep ni-s̄ewsew n̄o-töt̄i ḡivix̄ en.  
ART-fire AOR-hot~IPFV ART-tree Syzygium DEF  
'The fire was starting to *heat up/burn* the apple tree.'
- (MTP.20) mahi t̄e-s̄ew k̄p̄w̄ijiŋ nk̄.  
place FUT<sub>1</sub>-hot FUT<sub>2</sub> 2sg  
'You're going to feel hot.'  
[lit. The place will be hot (on) you]

This transitive use of the adjective is not attested with *m̄omjij* 'cold', nor does it seem to be found in other languages. It is lexically specific to the verb *sew* of Mwotlap.

Finally, a very common construction treats temperature terms, figuratively, as agentive subjects. Indeed, some adjectives can be derived into nouns, either by reduplication (§3.2) or zero-conversion, and become compatible with noun determiners. Thus the adjective *m̄omjij* converts into a noun *m̄omjij*, which is commonly prefixed with the article, yielding *n̄o-m̄omjij* '[N] cold, coldness'. The expression of cold vs hot sensation is syntactically parallel to all non-controlled sensations such as hunger, thirst, sickness, sleepiness etc.: they involve an SVO construction, in which an NP referring to the feeling is the subject of a verb 'do' or 'affect' (François 2005:136). In (21) from Mwotlap, the subject is *na-matmajȳe* 'sleepiness'; in (22), it is

the noun *no-mɔmɔjij* '[N] cold, coldness', itself the result of a zero-conversion from the adjective.

- (MTP.21) na-matmajγε m-ak no.  
ART-sleepiness PRF-do 1sg  
'I'm feeling sleepy.' [lit. *Sleepiness is doing me*]
- (MTP.22) no-mɔmɔjij m-ak ki.  
ART-coldness PRF-do 3sg  
'She's (feeling) cold.' [lit. *Coldness is doing her*]

The nominalisation of 'hot' in Mwotlap is slightly different from that of 'cold', as it involves a compound noun *mahi-sew* (lit. 'hot place') based on *mahi* 'place' [cf. (14), (20)].

- (MTP.23) na-mahi-sew m-ak ki.  
ART-place-hot PRF-do 3sg  
'She's (feeling) hot.' [lit. *Hot-place is doing her*]

The various case frames involving temperature predicates are summarised in *Table 3*.

*Table 3* – Case frames involving temperature predicates

ex.	Subject	Predicate	Object	frequency
(7)-(10)	theme	TEMP. PRED.	—	common
(14)-(15)	dummy N 'place'	TEMP. PRED.	—	common
(20)	theme ~ 'place'	TEMP. PRED.	<b>experiencer</b>	rare: only Mwotlap <i>sew</i> 'hot'
(22)-(23)	TEMP. NOUN	verb 'do, affect'	<b>experiencer</b>	common
(13)	<b>experiencer</b>	TEMP. PRED.	—	rare: only Mwerlap <i>malas</i> 'feel.cold'
(17)-(18)	<b>experiencer</b>	verb 'feel'	embedded temp. clause	average
(25)	<b>experiencer</b>	verb 'feel'	TEMP. NOUN	rare

## 5 Lexical semantics

This last section will explore the lexical contrasts and semantic extensions which characterise temperature terms.

After a look at etymologies, I will discuss the cases where more than two terms are found in the domain of 'cold'. Second, I will briefly discuss other terms belonging to the semantic domain of temperature. Finally, I will come back to the two basic terms for 'hot' and 'cold', and survey their semantic extensions.

### 5.1 Etymologies

Most of the forms in *Table 2* are cognate across languages of the area. Apart from the innovative form *misip* in Lakon, of unknown origin, most of the words for 'cold' reflect the Proto Oceanic etymon \**mari(d)ri(y)* (Ross, Pawley, Osmond 2003:217) with the same meaning, via a reduplicated form \**mamadridri*. Hiw and Lo-Toga reflect a dissimilated variant \**maladridri*, while Lemerig, Vera'a and Vurës reflect another form of dissimilation \**mamaxidri*. These forms illustrate sound change, and do not affect meaning.

The adjectives for 'hot' split up in two distinct cognate sets. Nine languages reflect an etymon \**tutunu*, while five reflect \**savu*. Both of these forms reflect historical innovations compared to the etymon \**panas* which has been reconstructed (Ross *et al.* 2003:217) for Proto Oceanic 'hot'. Interestingly, these two innovations make reference to cooking practices. Thus, \**tutunu* 'hot' results from the reduplication of \**tunu* 'roast (food) on embers or fire' (Clark 2009:205). As for \**savu*, it originally refers to the hot steam that is produced when steaming food in the earth oven (Clark 2009:176) – see §2.

### 5.2 Two words for 'cold'

Languages of northern Vanuatu normally have just one term for 'hot'. The only exception is Lemerig, which has *?il?in* and *vævæt*; the two words were described as synonyms, and my corpus doesn't suggest

any clear semantic contrast. The case is a bit different for the ‘cold’ domain, as the languages with two different terms (*Table 2*) seem to treat them differently. These are the object of the next pages.

### 5.2.1 SYNTACTIC DISTINCTIONS

Four languages have a dedicated noun for 'cold' or 'coldness', which is not derived from the adjective. These nouns reflect a Proto Oceanic etymon \*malaso(*y*), which is also reconstructed as a noun '[N] cold' (Ross, Pawley, Osmond 2003:218; Clark 2009:135).

When using a nominal construction such as (22) above, some languages can choose freely between two strategies, with no major difference in meaning. Thus Vures can use its noun *malas*:

- (VRS.24) o *malas* ma=<sup>n</sup>da n  
 ART coldness PRF=do 1sg  
 'I am cold ~ I have fever.'

...but the same language can also easily resort to the faculty to convert (i.e. zero-derive) adjectives into nouns, and thus nominalise its adjective *mamixin*:

- (VRS.24') ○ *mamixin* ma=<sup>n</sup>da n○  
 ART cold PRF=do 1sg  
 'I am cold ~ I have fever.'

The noun *malas* can be the object of a verb 'feel':

- |                                       |     |          |     |               |
|---------------------------------------|-----|----------|-----|---------------|
| (VRS.25)                              | na  | mō=rōj   | o   | <i>malas.</i> |
|                                       | 1sg | PRF=feel | ART | coldness      |
| 'I am cold.' [lit. 'I feel the cold'] |     |          |     |               |

Besides its reference to plain cold temperature, the reflexes of \**malasonj* colexify<sup>4</sup> 'cold' with 'fever' or 'malaria' – as in Vur  s (24), or in Lo-Toga (26):



<sup>4</sup> On the term *colexify*, i.e. ‘lexify (two or more meanings) with the same form’, see François (2008a).

The two cognate sets *\*mamadridri* and *\*malasoy* apparently target the same range of temperatures. Their only difference seems to be morphosyntactic, as they belong to two different parts of speech. Many languages have historically lost one of the two terms, presumably due to their semantic redundancy. As for the Mwerlap verb *malas* ‘feel cold’ illustrated in (13), it reflects a change of the root’s syntactic category from noun to verb – an innovation which is apparently specific to this language.

### 5.2.2 SEMANTIC DISTINCTIONS

Hiw is a different case: it has two distinct adjectives for 'cold', *məwmɪw* and *maetit*, with differences in meaning. A possible description of the Hiw terms would involve a contrast between 'subjectively cold' (*məwmɪw*) and 'objectively cold' (*maetit*).

The adjective *məwm̥iɪw* means ‘cold’ applying to the weather or to ambient temperature, but also to a body of water (sea, river) where someone bathes. The meaning here is the effect of the temperature on thermoregulation, whether in water or in air – that is, thermal comfort. The connotation of *məwm̥iɪw* can be positive (‘nicely cool’) or negative (‘freezing cold’) – an ambiguity which is sometimes solved by modifying adverbs.

- (HIW.27) nə ɲʷutə ne məwəmwıw. (\*ne maetit)  
 ART place STAT cold1  
 'It (the weather) is cold.'  
 [whether nicely cool or unpleasantly cold]

- (HIW.28) nə pɪgɔnə ne məwəmwiw tinniŋL. (\*ne maetit)  
 ART sea STAT cold1 appropriate  
 'The seawater is nicely cool.'

- |                                  |     |        |      |          |       |              |
|----------------------------------|-----|--------|------|----------|-------|--------------|
| (HIW.28')                        | nə  | pɪgɔnə | ne   | məwəmriw | tñiy. | (*ne maetit) |
|                                  | ART | sea    | STAT | cold1    | INTSF |              |
| 'The seawater is freezing cold.' |     |        |      |          |       |              |

Metonymically, *məwm̩iŋ* as an attribute with ‘clothes’ means ‘light clothes’, i.e. ‘clothes that make one feel cooler’:

- (HIW.29) nə jojə məwm̩i w (\*maetit)  
 ART clothes cold1  
 'cold/cool/summer clothes'

The opposite phrase *nə jojə ttin* would refer to ‘warm clothes’ such as woolly jumpers - items which are seldom seen in Vanuatu, but occasionally mentioned.

By contrast, *maetit* means ‘cold’ in the TACTILE sense, as an inherent property of an object. An example of its use is when food meant to be served hot “gets cold” [cf. (8), (10)]:

- (HIW.30) nə təg̩ov ne maetit pə. (\*ne məwm̩i w)  
 ART pudding STAT cold2 now  
 'The pudding is cold now (unsuitable for eating).'

The semantics of *məwm̩i w* thus pertains to the domains of AMBIENT and PERSONAL-FEELING TEMPERATURE, whereas *maetit* is TACTILE. This may recall the contrast between, respectively, Japanese *samu-i* ‘cold [AMBIENT]’ and *tsumeta-i* ‘cold [TACTILE]’. However, the match is not perfect, because Japanese *samu-i*, which could indeed translate *məwm̩i w* in (27), would be excluded from sentences like (28) or (29).

While they can be contrasted, the two adjectives of Hiw occur in free variation in a specific context: namely, when a food or drink is so cold that its consumption brings a cooling sensation, whether pleasant or not, to the whole body:

- (HIW.31) nəkə ni nə mətu, ne maetit ~ məwm̩i w tate<sup>q</sup>lītə.  
 1sg drink ART coconut STAT cold2 cold1 suitable  
 'I'm drinking a coconut, it is nicely cool.'

The two adjectives are not exactly synonymous here: while *maetit* describes the coconut juice as being objectively cold, *məwm̩i w* describes its effect on the experiencer, in a way similar to Eng. *refreshing*.

### 5.3 Other terms related to temperature

In sum, the languages of north Vanuatu have no more than two or three words that can, strictly speaking, be called temperature terms.

In most cases, this involves a binary contrast between ‘hot’ and ‘cold’, with no further terms on the thermal scale. Worthy of notice, for example, is the absence of any term in the middle, equivalent to Eng. *tepid* or *lukewarm*. Likewise, these languages have no lexical term to describe extreme temperatures (like *freezing* or *burning*) or nuances within the cold or hot domains (as Eng. *cool* vs *cold*, or *warm* vs *hot*).

Rather than using other adjectives that would be in the same paradigm as the basic adjectives ‘hot’ and ‘cold’, most nuances on the scale are expressed by adverbs - i.e. modifiers of the predicate head. This is true of some moderating adverbs, such as Hiw *tinni<sup>q</sup>l* ‘appropriately’ in (28), *tate<sup>q</sup>lītə* ‘suitably’ in (31), or Mwotlap *tusu* ‘a bit’ in (10). This is also true of intensifiers, such as Mwotlap *meh* ‘too much’ in (7), or Hiw *tñiq* in (15) or (28). As we saw in §3.2, reduplication alone can have an intensifying function - e.g. *sewsew* ‘very hot’ in (33) below. In addition, Mwotlap has lexically specific intensifiers for a large number of adjectives (François 2001:266). The one for ‘cold’ is (*mɔmij*) *"beləwət*, of unknown origin - it would be used to describe, for instance, some *very* cold water. The intensifier for ‘hot’ is (*sew*) *evej* - which typically refers to red-hot (incandescent) stones in the earth oven, as in (9) above. Mwotlap *evej* is cognate with Mota *vevera* ‘red-hot’, and ultimately with the root \**we-wela* ‘hot’ of Proto-Polynesian (>Tahitian *vera* ‘hot’, etc.).

Also indirectly related to the temperature domain would be a verb like Mwotlap *s(ɔ)k̩p<sup>w</sup>ɔ* ‘scald’. The body part affected is either the subject of the verb used intransitively - as in (32) - or the object of a transitive construction, as in (33):

- (MTP.32) na-mn̩i-k m̩e-sk̩p<sup>w</sup>ɔ.  
 ART-hand-1sg PRF-scald  
 'I scalded my hands (with hot water).'

- (MTP.33) ni<sup>m</sup>b̩i sewsew m̩e-sk̩p<sup>w</sup>ɔ na-χajm̩e-k.  
 ART-water hot<sup>m</sup>INTSF PRF-burn ART-tongue-1sg  
 'I scalded my tongue with the tea.'  
 [lit. 'the hot water scalded my tongue']

Possibly related is the verb *s(ε)k̪ʷεŋ* ‘burn (skin) due to exposure to heat or sun’. The neighbouring language Mota has cognate verbs *sak̪ʷo* ‘burn with heat, provoke a sensation of burning’ and *sak̪ʷora* ‘scorch’ (Codrington & Palmer 1896:160).

Finally, the languages of the region also have rich lexicons in the domains of fire (light a fire; burn; glow; smoulder; flames; sparks; move fire from one place to another; slash-and-burn garden...) or of cooking (cook; fry; roast; bake; steam; manipulate hot stones; be cooked; etc.). These words, however, form a separate domain from the set of temperature words strictly speaking, and I will not detail them further. The same could be said of weather terms (rain; drizzle; haze; drought; wind; hurricane...) or other words which are only indirectly related to temperature (sweat; shiver; shade; etc.).

## 5.4 Semantic extensions of HOT and COLD

### 5.4.1 GENERAL COMMENT

Let us now come back to the two adjectives ‘hot’ and ‘cold’, which are the only two (or three) temperature terms, strictly speaking, found in northern Vanuatu languages. As we saw in various examples, the reference of these two adjectives is both to TACTILE and to AMBIENT temperature. One may include PERSONAL-FEELING temperature in their semantic content, except that, with the exception of (13) above, these terms are hardly ever used as predicates taking the experiencer as their subject (§4.2.3).

It would be pointless to define the semantic difference between HOT and COLD in terms of absolute temperature, since it is obviously bound to the context. Water at 26°C would be described as ‘warm/hot’ or ‘cool/cold’ depending on the outside temperature or the physical condition of the experiencer. Also context-dependent are the evaluative connotations which would be carried by these terms: neither ‘hot’ or ‘cold’ can be said to be intrinsically meliorative or pejorative in these languages. Simply, given the tropical weather of Vanuatu and its typically high temperatures, it is more frequent to

hear people complain about the weather being ‘hot’ than the opposite; in such a context, the use of ‘cool/cold’ would have positive connotations – as in this Mwotlap example:

- (MTP.34) *ɣm van hay mahi mɔmjiŋ!*  
1inc:pl go sit place cold  
'Let's go sit somewhere cold/cool!' (i.e. shady)

As mentioned in the introduction, the two temperature adjectives display little in way of semantic extension – especially in comparison with European languages, in which this domain is fertile in polysemies. Only two or three senses, other than strict temperature meanings, can be reported for each adjective of the pair.

### 5.4.2 FEVER

First, a PERSONAL-FEELING predicate equivalent of Eng. *X is (feeling) cold* – as in (22) above – may refer not only to the feeling of temperature strictly speaking, but also, by extension, to the sensation of intense shivering that accompanies fever. In the context of tropical Vanuatu, such a condition is basically synonymous with malaria:

- (MTP.35) *nɔ-mɔmjiŋ me-wsɛŋ ki.*  
ART-coldness PRF-shake 3sg  
'She has a strong (malaria) fever.'  
[lit. *Coldness is shaking her*]

In Mwotlap, this sort of feverish cold is sometimes distinguished from mere temperature statements by means of a compound *nɔ-mɔmjiŋ ɣɔm*, literally ‘coldness<sub>[N]</sub> of disease’:

- (MTP.36) *nɔ-mɔmjiŋ ɣɔm m-ak ki.*  
ART-coldness disease PRF-do 3sg  
'She has (malaria) fever.'

Löyöp uses the reverse phrase *n-ɣɔm mam⁹din*, lit. ‘disease of coldness’:

- (LYP.37) *n-ɣɔm mam⁹din m-an nø.*  
ART-disease coldness PRF-do 1sg  
'I have (malaria) fever.'

The language Lakon would express the same meaning with a similar phrase *misiŋ ya lekteŋ*, literally ‘strong coldness<sub>[N]</sub>’:

- (LKN.38) misiŋ ya lekteŋ m tʃiŋ na.  
coldness STAT strong PRF affect 1sg  
'I have (malaria) fever.'

The notion of fever can be expressed via both the domain of COLD – due to the experience of the shivering person – and that of HOT – due to the external manifestation of fever for an outsider. Thus, ‘be feverish’ in Mwotlap is a form *kpwitsew*, itself a compound of *kpwit-* ‘head’ and *sew* ‘hot’:

- (MTP.39) titi nənəm ki mu-su kpwit-sew.  
baby your 3sg PRF-DIMIN head-hot  
'Your baby is a bit feverish.' [lit. *a little hot-headed*]

Logical though it may be, this flip-flop between ‘hot’ and ‘cold’ may occasionally sound odd when the two domains are being used in the same sentence – as in (40), taken from a story:

- (MTP.40) nə-mɔmjiŋ yɔm n-ak no a uu!  
ART-coldness disease STAT-do 1sg INTSF EXCL  
takpwitm-wr-k sewsew en!  
body-1sg hot~INTSF EXCL  
'Hey, I'm having a terrible fever, [lit. *a terrible COLD*]  
my body's burning HOT!'

#### 5.4.3 SCORCHING SENSATIONS

The term ‘hot’ – e.g. Mwotlap *sew* – also has other semantic extensions. Besides the temperature meaning, one common use is for spicy food:

- (MTP.41) ne=m̥bep yoh ne-sewsew, aj!  
ART-chilli this STAT-hot~INTSF EXCL  
'Oh my god, this chilli is super hot!'

The same word would be used for the physical sensation due to a burn or stinging pain, e.g. after an insect bite or contact with a stinging nettle tree (*Dendrocnide* sp.). Note the transitive use of the adjective here (cf. (20)):

- (MTP.42) no ma-taŋ van na-hlat, me-sew na-mni-k!  
1sg PRF-touch thither ART-nettle PRF-hot ART-hand-1sg  
'I touched a nettle tree, my hand's *burning!*'  
[lit. 'it's heating my hand']

In Mota, such senses are expressed with *sakpwō* ‘provoke a burning sensation’ mentioned in §5.3.

#### 5.4.4 PSYCHOLOGICAL SENSES

I have found three psychological extensions of ‘hot’, which are all rare in spontaneous speech.

One is the Mwotlap compound verb *ak sewsew* (or *yaleŋ sewsew*) literally ‘make hot, warm s.th. up’. It is sometimes used to mean ‘encourage, cheer up s.o.’:

- (MTP.43) ym sc ak sewsew iŋε "ba-laklak.  
1inc:pl PROSP do hot~RESULT HUM:PL for-dance  
'Let's go cheer up the dancers.' [lit. 'warm them up']

- (MTP.44) nɔ-hɔhɔlɛ nənəm ma-yaleŋ sewsew nɔ.  
ART-speech your PRF-make hot~RESULT 1sg  
'Your words have comforted me / given me courage.'

The other compound is from the language Vurës: “dø=døm tütün”, literally ‘think hot’. It refers to the state of mind of someone who is focused and determined to achieve their goal:<sup>5</sup>

- (VRS.45) na yœvrü-n rɔrɔ inε, si rɔrɔ mɔ="dø=døm tütün  
ART house-POSS 3du this if 3du PRF=think hot  
ai, timiak inε ni mi=kpwit ren ti.  
OBL:ADV like this 3sg PRF=finish long.ago CTFC  
'That house of theirs, if only they had been serious  
[lit. *thought hot*] about it, at this stage it would have been  
finished long ago.'

The parallel phrase in Mota is *nom tutun* ('think hot'), which Codrington’s dictionary glosses as ‘to be earnest’ (Codrington &

<sup>5</sup> The sentence is from Hyslop-Malau (n.d.); the transcription, glosses and translation are mine.

Palmer 1896:221). Yet these examples are isolated: I have not found any other similar use of ‘hot’ elsewhere in northern Vanuatu.

Finally, ‘hot’ is sometimes associated with *anger* – as though one’s inner self were being burnt by the feeling of anger. While the metaphor is common in languages like English (Lakoff & Kövecses 1987), in Vanuatu it is relatively rare. As far as the Banks-Torres languages are concerned, the only case I am aware of is found in the dictionary of Mota (Codrington & Palmer 1896:160):

(MTA.46)	na	lolo-na	we	sakp <sup>w</sup> o.
	ART	inside-3sg	PRST	scorch

‘He is hot with anger.’ [lit. ‘his mind is scorching.’]

It is common for northern Vanuatu languages to express strong feelings like anger with idioms or compounds derived from the root \**lolo* ‘inside, inner mind’ (François 2013:205 sqq.). Depending on the language, *anger* will translate as ‘[mind] bad’, ‘[mind] biting’, ‘[mind] itchy’ or ‘[mind] blowing strongly (like a hurricane)’. However, the Mota sentence in (46) is the only example in which anger is specifically linked to the semantic domain of heat.

This being said, the same semantic link is found in some related languages outside the focus area of this study. In Araki, a moribund language spoken further south in Vanuatu, the verb *lolokoru* ‘be angry’ is a compound of *lolo* ‘inside, inner mind’ and *koru* ‘dry, desiccated, burnt by fire’ (François 2002:265, 2008b).

Beyond their differences, the three psychological meanings have in common the association of HOT with energy and resolve in performing an action. The polysemies do not form a symmetrical system: that is, the opposite meanings – calm, weakness or laziness – are never explicitly associated with the domain of ‘cold’.

## 5.5 A relative poverty in semantic extensions

No other significant semantic extension of HOT and COLD has been found in these Vanuatu languages, beyond the few and isolated cases

just described. This is not just an artefact of limited data, as can be judged by the size of the text corpus, as well as the author’s in-depth exposure to spoken conversation for long periods since 1997. Besides, a survey of various dictionaries from other Oceanic languages (whether from elsewhere in Vanuatu, or from New Caledonia, Polynesia, etc.) usually yields terse definitions for ‘hot’ and ‘cold’, with little polysemy reported anywhere.

This virtual absence of semantic extension for temperature terms deserves to be noted, as it stands in contrast with various other languages of the world for which such terms commonly integrate psychological or evaluative meanings, as described in other chapters of this volume. Besides, the contrast can be drawn also language-internally, by highlighting seeing how these languages of Vanuatu have developed rich polysemies in various domains other than temperature. For example, whereas the contrast *hot-cold* remains mostly tied to its literal meaning, a pair of adjectives such as *heavy* vs *light* commonly takes over a wide array of metaphorical senses. The following is an excerpt from my Mwotlap dictionary (François, in prep.) for the adjective *dēw* ‘heavy’:

**dēw** [ndi<sup>w</sup>] ADJ. heavy. *Ant.* momya  
'light'.

(1) heavy, weighty. ▷ **Ne-vet gōh nē-dēw meh.** *This rock is too heavy.*

(2) causing a feeling of fatigue or numbness, similar to the effects of carrying a heavy weight. ▷ **Nēk yoñteg na-taybē ni-dēw aē, tō nēk mitiy galsi ainqōn.** *(narcotic effects of kava beverage) You feel your body become heavy, which allows you to sleep well at night.*

(3) (metph) inspiring awe and respect, espec. in the context of traditional society; hence imposing, grave, serious, solemn. ▷ **Ige mōlmōl van la-halgyoy a kem et lēs tamat gete: nē-dēw mi kemem, kem nē-dēmap so kem so hayeg van.** *When men used to gather together in their secret societies, we the non-initiated, were*

*quite impressed [lit. it was heavy for us]: we felt too much awe to join in. Syn. map.*

(4) (topic, story+) important, major, serious; of outstanding social or psychological significance. ▷ **Nē-dēmdēm nan nē-dēw a nē-dēw.** *This is a very important topic.* ▷ **Na-kaka t-añag gōh nē-dēw lē-vēnangēn.** *This myth is a major one in our culture. Syn. liwo 'big'.*

(5) (artwork, style+) of great aesthetic value, of impressive beauty (vs. momya ‘lightweight, careless, slapdash’): grand, superb, awesome, magnificent, elegant, noble. ▷ **Nok so mōk na-plakas van l-eh so kē n-ak dēw na-he.** *I wanted to add a few nice turns to the poem in order to make your ode even more majestic [lit. to make it heavier].*

- (6) accomplished, refined, elaborate, tasteful; artfully performed. ▷ *Ige lōmlōmgep nen kēy laklak namuy en, et dēw galsi te.* These young people have a way of dancing that is not very graceful [lit. not very heavy].
- (7) (word, phrase) well phrased, elegant, refined; of a formal style, imbued with respect (vs *vasawyeg*

'informal, casual'); hence idiomatic, proper. ▷ "Heleg" nē-dēw *veteq "vēlēs"*. "Heleg" is a more formal word than "vēlēs". ▷ *Itōk, ba et dēw galsi te.* (This sentence) is correct, but it's not very idiomatic [lit. not heavy enough]. Syn. **mutuw** 'idiomatic'.

The elaborate polysemies found around the notion *heavy* contrast with the relative poverty of metaphorical extensions surrounding temperature terms. The following two entries illustrate again this point; even though they only exemplify one language, Mwotlap, they can be taken as a final synthesis for this whole case study.

### **momyiy** [mɔmɔjɪj] [ADJ]

- (1) (s.th.) cool, cold in temperature. ▷ **nē-bē momiy** cold water ▷ *Mahē no-momyiy*, (ambient) It's cold. Ant. **sew**.
- (2) (esp) (hot food) become cold, cool down. ▷ **Nēk ukēg kē ni-momyiy**. Leave it to cool.
- ◊ **[N] (no-momyiy)** (1) cold, feeling of cold. ▷ *No-momyiy m-ak no, nok so gom*. I'm cold [lit. Cold is doing me], I'm going to be ill. ▷ *Awēē, no-momyiy ee!* Oh my god, what a cold! Ant. **mahē-sew**.
- (2) (hence) sensation of cold due to fever; fever. ▷ **No-momyiy ni-wseg no**. [the cold is shaking me] I'm shivering ~ I have fever.

- ♦ **no-momyiy gom** N. lit. "sick coldness": malaria; disease characterised by an intense feeling of cold. ▷ **No-momyiy gom n-ak no a uu!** Oh, I've got terrible malaria! Cf. **gom** 'sick'.

[< \**mama'riri*; POC \**ma-(d)ridrig* 'cold']

### **SEW** [səw] [ADJ]

- Redup. sewsew.* (1) (s.th.) warm, hot in temperature. ▷ **Ne-gengen ne-sew leptō**. The food is still hot. ▷ **Egoj n-ep, veg ne-sew**. Beware the fire, it's hot. ▷ **Ōw ni-tintin hōw nen e, ne-sew a ne-sew evey!** He took out the wild yams (from the fire), they were burning hot! ▷ **Na-lo ne-sew, ay!** The sun's so hot! ▷ **Mahē ne-sew meh.** (ambient) It's too hot here. Intsf. evey.

- ♦ **nē-bē sewsew** N. lit. "hot water": hot drink, a Western tradition now popular in the islands: *gen. tea, somet. coffee, hot chocolate*. ▷ **Nē-bē sewsew tiple soqo na-nye!** Don't burn yourself with your tea! Syn. **ti**; See **bē**.
- (2) (rare) hot due to fever. ▷ **Taqmē-k sewsew en!** (fever) My body's burning hot! See **momyiy** 'cold'.
- (3) [resultative serialisation; always reduplicated] (do V) so as to render (s.th., s.o.) hot; warm up, heat up. ▷ **Nok so mihil sewsew no van l-ep**. I want to warm myself up (by standing) close to the fire.

### ♦ **ak sewsew ~ galeg sewsew** VT. (metph) lit. "make hot": warm up (s.o.) with one's words; cheer up, encourage, comfort.

▷ **No-hohole nōnōm m-ak sewsew no**. Your words have comforted me.

- (4) (chilli+) pungent, spicy. ▷ **Ne-bep gōh ne-sewsew, ay!** This chilli is terribly hot!

◊ [VT] (1) (s.th.) bring heat to (s.th., s.o.), overheat, burn. ▷ **N-ep ni-sew-**

**sew nō-tōti gēvēg en.** The fire was starting to catch at the apple tree.

▷ **Na-lo me-sew no.** I'm being burnt by the sun. Cf. **soqo** 'scald', **seqēn** 'burn'.

- (2) (plant, medicine+) burn, irritate (body part). ▷ **Na-hlat me-sew na-yñē-k**. My leg is burning due to (contact with) a stinging nettle. Syn. **gaygay**.

[< \**sawu*; PNCV \**savu(a)* 'steam, moist heat']

## Abbreviations

Abbreviations for language names include:

HIW: Hiw; LKN: Lakon; LTG: Lo-Toga; LYP: Löyöp; MRL: Mwerlap; MTA: Mota; MTP: Mwotlap; VRS: Vurës; POC: Proto Oceanic.

Example sentences are glossed according to the Leipzig rules. More specific abbreviations are listed here.

ADV	adverb	DU	dual	OBL	oblique
AOR	aorist	INTSF	intensifier	POSS	possessive
APPREH	apprehensive mood	COORD	coordinator	PRF	perfect
ART	article	CPLT	complete	PROSP	prospective
CTFC	counterfactual	HUM	human article	RESULT	resultative
DEIX	deictic	IPFV	imperfective	STAT	stative
EXCL	exclamative	IRR	irrealis		

## References

- Clark, Ross. 2009. *\*Leo Tuai: A comparative lexical study of North and Central Vanuatu languages*. Pacific Linguistics, 603. Canberra: Australian National University.
- Codrington, Robert H., & John Palmer. 1896. *A dictionary of the language of Mota, Sugarloaf Island, Banks Islands*. London: Society for Promoting Christian Knowledge.
- Crowley, Terry. 2000. The language situation in Vanuatu. *Current Issues in Language Planning* 1 (1):47-132.

- François, Alexandre. 2000. Vowel shifting and cloning in Motlav: historical explanation vs. formal description. In *Proceedings of AFLA 7 (The Seventh Meeting of Austronesian Formal Linguistics Association)*, Marian Klamer (ed.), 49-68. Vrije Universiteit Amsterdam: Amsterdam.
- . 2001. Contraintes de structures et liberté dans l'organisation du discours. Une description du mwotlap, langue océanienne du Vanuatu. Doctoral dissertation, Université Paris-IV Sorbonne, Paris.
- . 2002. *Araki. A disappearing language of Vanuatu*. Pacific Linguistics, 522. Canberra: Australian National University.
- . 2003. *La sémantique du prédicat en mwotlap (Vanuatu)*. Collection Linguistique de la Société de Linguistique de Paris, 84. Paris, Louvain: Peeters.
- . 2004a. La réduplication en mwotlap : les paradoxes du fractionnement. In *Les langues austronésiennes [Faits de langues n°23-24]*, E. Zeitoun (ed.), 177-194. Paris: Ophrys.
- . 2004b. Chains of freedom: Constraints and creativity in the macro-verb strategies of Mwotlap. In *Complex predicates in Oceanic languages: Studies in the dynamics of binding and boundness [Empirical Approaches to Language Typology]*, I. Bril and F. Ozanne-Rivierre (eds), 107-143. Berlin: Mouton de Gruyter.
- . 2005. A typological overview of Mwotlap, an Oceanic language of Vanuatu. *Linguistic Typology* 9-1:115-146.
- . 2006. Serial verb constructions in Mwotlap. In *Serial Verb Constructions: A cross-linguistic typology [Explorations in Linguistic Typology]*, R.M.W. Dixon & A. Aikhenvald (eds), 223-238. Oxford: Oxford University Press.
- . 2008a. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In *From Polysemy to Semantic change: Towards a Typology of Lexical Semantic Associations [Studies in Language Companion Series 106]*, Martine Vanhove (ed.), 163-215. New York, Amsterdam: Benjamins.
- . 2008b. An online Araki-English-French dictionary. Electronic publication: <<http://alex.francois.free.fr/AF-Araki.htm>> (accessed 8/04/2013)..
- . 2011a. Where \*R they all? The geography and history of \*R loss in Southern Oceanic languages. *Oceanic Linguistics* 50 (1): 140-197.
- . 2011b. Social ecology and language history in the northern Vanuatu linkage: A tale of divergence and convergence. *Journal of Historical Linguistics* 1 (2): 175-246.
- . 2013. Shadows of bygone lives: The histories of spiritual words in northern Vanuatu. In *Lexical and structural etymology: Beyond word histories [Studies in Language Change]*, R. Mailhammer (ed.), 185-244. Berlin: DeGruyter Mouton.
- . in prep. A Mwotlap-French-English dictionary.
- Hyslop-Malau, Catriona. n.d. Dictionary of Vurës. Electronic publication: <<http://paradisec.org.au/vandicts/index.htm>> (accessed 8/04/2013).
- Koptjevska-Tamm, Maria. 2011. "It's boiling hot!": On the structure of the linguistic temperature domain across languages. In *Rahmen des Sprechens. Beiträge zur Valenztheorie, Varietätenlinguistik, Kognitiven und Historischen Semantik*, Sarah Dessì Schmid, Ulrich Detges, Paul Gévaudan, Wiltrud Mihatsch and Richard Waltereit (eds.), 379-396. Tübingen: Narr.
- Lakoff, George & Zoltán Kövecses. 1987. The cognitive model of anger inherent in American English. In *Cultural models in language and thought*, D. Holland & N. Quinn (eds), 195-221. Cambridge: Cambridge University Press.
- Launey, Michel. 1994. Une grammaire omniprédictive: Essai sur la morpho-syntaxe du nahuatl classique. Sciences du Langage. Paris: CNRS.
- Lemaréchal, Alain. 1989. *Les parties du discours, Syntaxe et sémantique. Linguistique Nouvelle*. Paris: Presses Universitaires de France.
- Lynch, John, & Terry Crowley. 2001. *Languages of Vanuatu: A new survey and bibliography*. Pacific Linguistics, 517. Canberra: Australian National University.
- Ross, Malcolm, Andrew Pawley, & Meredith Osmond, eds. 2003. *The lexicon of Proto Oceanic: The physical environment*. Pacific Linguistics, 545. Canberra: Australian National University.
- Tryon, Darrell. 1976. *New Hebrides Languages: An internal classification*. Pacific Linguistics, C-50. Canberra: Australian National University.

Sébastien Lacrampe  
Linguistics  
School of Culture, History and Language  
Australian National University  
Canberra  
Australia

Canberra, 12 February 2014

TO WHOM IT MAY CONCERN

I write as one of the editors of the following volume:

François, Alexandre; Sébastien Lacrampe; Stefan Schnell & Michael Franjieh (eds),  
***The Languages of Vanuatu: Unity and Diversity.*** Studies in the Languages of  
Island Melanesia, 4. Canberra: Asia Pacific Linguistics Open Access.

This edited volume is currently in the making, and should be released soon  
(hopefully this year 2014) in *Asia-Pacific Linguistics*' new academic series 'Studies  
in the Languages of Island Melanesia'

(<http://pacling.anu.edu.au/materials/Studies-in-the-Languages-of-Island-Melanesia.htm>)

While Dr Alexandre François is indeed one of the editors of the forthcoming volume,  
he is also the author of one of its chapters, entitled:

"The ins and outs of *up* and *down*: Disentangling the nine geocentric space systems  
of Torres and Banks languages"

I have been the editor in charge of having this article going through the double-blind  
peer-reviewing process, as per the requirements of the series. This letter is to certify  
that the reviews we received for A. François' article have motivated us to accept it in  
our volume, pending minor revisions which are due in mid March 2014.

Thanks for your attention,



Sébastien LACRAMPE



# The ins and outs of *up* and *down*

## Disentangling the nine geocentric space systems of Torres and Banks languages

Alexandre FRANÇOIS

LACITO-CNRS; Australian National University

### Abstract

The 17 Oceanic languages spoken in the Torres and Banks Islands of northern Vanuatu commonly encode spatial relations by resorting to an absolute “geocentric” system of space reference. These systems all have in common a single cardinal axis oriented northwest-southeast, and at least a second topographical axis, contrasting inland-seawards. But while this general profile is typical of Oceanic, a detailed comparison of the 17 languages reveals their unexpected diversity, with nine distinct geocentric systems represented in this small region. The aim of this study is to describe and analyse these nine systems, through the systematic examination of the paradigms of space directionals that encode them. By mapping all the vectors involved onto an etic grid, I outline and compare each language's emic categories. Based on these empirical observations, I then propose a unified theory to account for the similarities and differences among these space systems, and to reconstruct their historical development since Proto Oceanic.

### Citation

François, Alexandre. (f/c). The ins and outs of up and down: Disentangling the nine geocentric space systems of Torres and Banks languages. In *The Languages of Vanuatu: Unity and Diversity*, ed. by A. François; S. Lacrampe; S. Schnell & M. Franjieh. Studies in the Languages of Island Melanesia. Canberra: Asia Pacific Linguistics Open Access.

## 1 The geocentric use of directionals

### 1.1 Space strategies across languages

The array of devices used by languages to construct their reference includes spatial expressions – that is, linguistic forms whose main function is to encode a direction or a location in the three-dimensional space. Studies of space systems used throughout the world have shown the existence of substantial variation across languages in the strategies they use for defining their coordinates. Figure 1 summarises the typology of linguistic space strategies as outlined by Levinson (1996b:359), based on the way they construct systems of spatial coordinates.

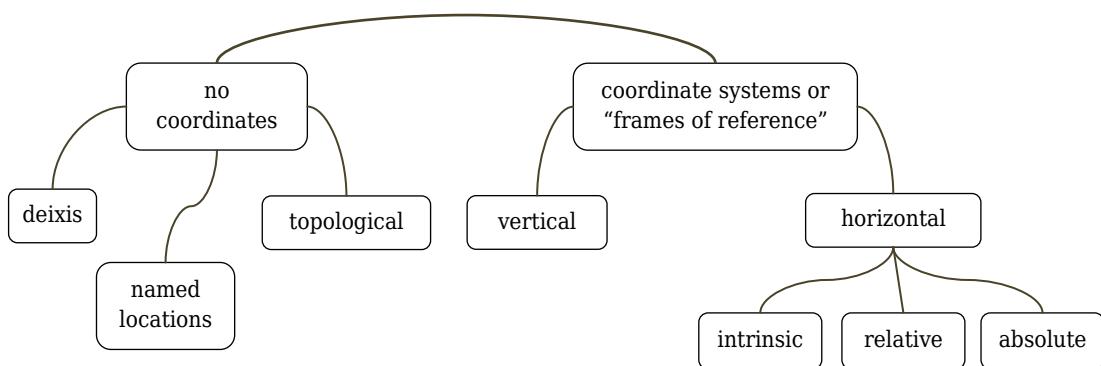


Figure 1 — A typology of space strategies (after Levinson 1996b:359)

While deictic (e.g. *here*) or topological strategies (e.g. *in the box*; *close to the tree*), as well as vertical coordinates, appear to be encoded in all languages, considerable cross-linguistic variation is observed regarding the strategies used on the horizontal plane. As Levinson (1996b, 2003) shows, languages can encode the relevant vectors by potentially resorting to three frames of reference. The INTRINSIC frame bases coordinates on the intrinsical orientation of a ground object taken as reference (Eng. *the ball is in front of the house*); the RELATIVE frame defines them with respect to the relative orientation of a ground object and a human observer (Eng. *the ball is in front of the tree*); finally, the ABSOLUTE frame of reference encodes directions based on an abstract system of fixed coordinates that do not depend on any particular anchor in the speech situation (Eng. *my house is south of the hill*). Crucially, several studies (Brown & Levinson 1992, 1993; Haviland 1993; Levinson 1996a-b, 2003; Pederson *et al.* 1998) have pointed out that these different frames of reference are diversely represented in the world's languages. English resorts to the three strategies, but uses the absolute frame only for longer distances (e.g. *north of England*, *western suburbs*). By contrast, a language like Tenejapan Tzeltal in Mexico (Brown & Levinson 1992) resorts exclusively to the absolute frame of reference, including for short distances; whereas Mopan Maya relies heavily on the intrinsic strategy (Pederson *et al.* 1998:572; Danziger 2011).

In the Oceanic family that is central to this study, most languages encode their horizontal directions by resorting to the *absolute* strategy (Palmer 2002). Unlike European languages which restrict their cardinal directions to long distances, Oceanic languages employ their absolute systems for any distance, and it is common to hear sentences such as this one in Mwotlap (François 2003:420):

(MTP.1)	na-bankēn	mey	hag	tō	lok	<b>hag</b>
	ART-mug	REL	sit	PRSTV	side	east
'the mug which is <i>on the east side</i> (of the table)'						

In the islands of northern Vanuatu which will be the focus of this study, absolute coordinates are in fact the only strategy possible for horizontal directions. Even though there are words for *right* and *left* referring to sides of the body, these words are never used to encode any spatial domain. Likewise, equivalents of *in front* or *behind* are never used for spatial reference – whether in an INTRINSIC frame (*behind the house*) or a RELATIVE one (*behind the tree*). If a figure X is located at a place that an English speaker would describe as ‘behind (Y)’, the only acceptable strategy in northern Vanuatu languages, in order to locate X with respect to Y, is to resort to ABSOLUTE coordinates (Hyslop 2002:74). This yields such sentences as (2), also in Mwotlap:

(MTP.2)	Kē	mi-tig	lō-tōti	beg,	ba	lok	<b>hōw.</b>
	3sg	PRF-stand	LOC-trunk	breadfruit	but	side	west
(liter.) ‘She’s standing at the breadfruit tree, on the western side.’							

Various recent publications have described the space systems of Oceanic languages: see Ozanne-Rivierre (1999), Cablitz (2006), Palmer (2007), as well as papers in Senft (1997), Bennardo (2002). Descriptions of space systems in languages of Vanuatu have been few so far, but include Hyslop (2002) on Northeast Ambae; François (2003) on Mwotlap; Paviour-Smith (2009) on Aulua. Except for Mwotlap, the space systems presented in this study are all described here for the first time.

## 1.2 Geocentric space reference in northern Vanuatu languages

In the Oceanic languages of northern Vanuatu, absolute systems of space reference thus form the default strategy for encoding any horizontal direction, regardless of distance. Not surprisingly, they are pervasive in discourse, both in narratives and in daily conversation. Whether they express motion paths or point to static locations (§2.3), these systems of absolute coordinates form the core of spatial reference.

The present study proposes to describe the various systems of absolute reference used in the seventeen languages of the Torres and Banks Islands, in the northernmost part of the Vanuatu archipelago. These languages do not just employ a simple system of fixed cardinal directions such as *north-west-south-east*: instead, they use paradigms of 3, 4, 5 or 6 directionals, whose defining parameters are based on several salient features of the islands’ geography: sea, shore, forest, winds and shape of islands... Due to this strong affinity with the geographical features of the landscape, I will describe these absolute systems with the term GEOCENTRIC (cf. Dasen & Mishra 2010).

I will focus on one syntactic category that is found in all the languages under study, namely *space directionals* (§2.3), and show that the vectors encoded by these directional paradigms are of three types:

- PARTICIPANT-oriented coordinates,  
glossed ‘hither’–‘thither’ [→§2.4.1];
- TOPOLOGICAL coordinates,  
e.g. ‘in’–‘out’, ‘up’–‘down’ [→§2.4.2];
- GEOCENTRIC coordinates,  
e.g. ‘inland’–‘seawards’, ‘uphill’–‘downhill’, ‘southeast’–‘northwest’... [→§2.4.3]

In the languages of northern Vanuatu, geocentric reference is almost always performed by directional particles which also have other, non-geocentric meanings. The typical case is that a given directional encodes topological coordinates as well as geocentric ones, following non-trivial patterns of correspondences. For example, all languages express NORTHWEST as DOWN; some languages encode SEWARDS as DOWN, others as OUT; and so on. Even though all northern Vanuatu systems share a number of general properties, the attested combinations result in relatively diverse systems of directionals.

In order to get a sense of this local diversity of geocentric systems, the paradigm of directionals in Dorig (Table 1) can be compared with the one in Hiw (Table 2).<sup>1</sup>

*Table 1 — The directional system of Dorig (Gaua island)*

Directional	PARTICIPANT-ORIENTED	TOPOLOGICAL	GEOCENTRIC
<b>ma</b>	hither	—	—
<b>āt</b>	thither	—	—
<b>vak</b>	—	across	parallel to shore in any direction, for short distances
<b>sag</b>	—	up; in	landwards, inland, uphill; (long-distance) parallel to shore towards SE
<b>ror</b>	—	down; out	seawards, downhill; parallel to shore towards NW

*Table 2 — The directional system of Hiw (Torres islands)*

Directional	PARTICIPANT-ORIENTED	TOPOLOGICAL	GEOCENTRIC
<b>me</b>	hither	—	—
<b>vēn</b>	thither	—	(on land) parallel to shore towards SE
<b>ag</b>	—	—	landwards, inland; (navigational) towards SE
<b>iy</b>	—	in	—
<b>rōw</b>	—	out	seawards
<b>vēn</b>	—	up	uphill
<b>uw</b>	—	down	downhill; (any distance) towards NW

### 1.3 The need to interpret modern directional systems

The geocentric directions can be plotted on a figure representing an island. In order to make systems comparable, I choose to represent all systems based on a single representation of the typical island landscape (§2.1, 2.4.3), to which each language is meant to adapt. Figure 2 maps the geocentric directionals of Dorig (Table 1); Figure 3 those of Hiw (Table 2). Whenever possible, each vector is tagged with a gloss representing its non-geocentric meaning: this will set the scene for the main point of this study, which is to understand the logics behind the lexification of geocentric directions in these languages.

The details of each language will be discussed later in this study. At this stage, the point is to observe that while all geocentric systems are similar in their general structure – e.g. through their orientation NW-SE, their distinctions between different scales, etc. – they can differ substantially in the fine-grained detail of their make-up. The differences are most

<sup>1</sup> Throughout this study, forms will be given using the local orthographies. A key to spelling and pronunciation is given in the Appendix (§11.2).

conspicuous in the way in which each language *lexifies* its directions, i.e. what lexical material it employs to encode them.

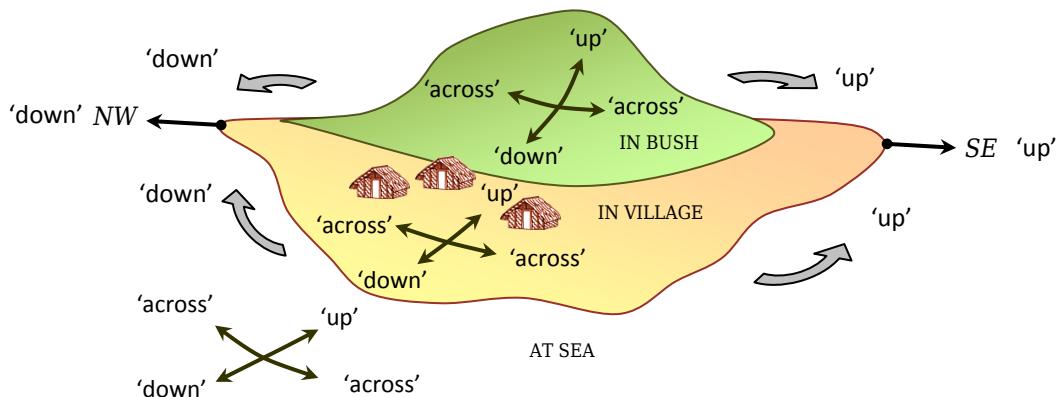


Figure 2 – The system of geocentric directionals in Dorig and other Gaua languages

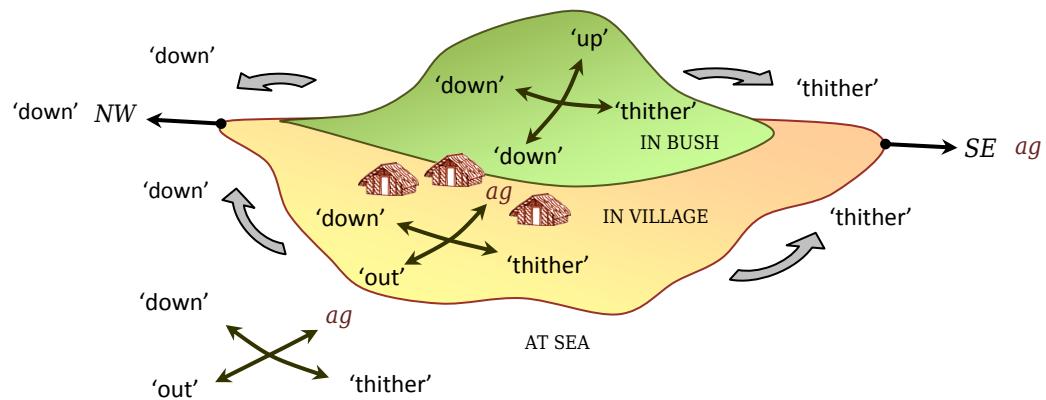


Figure 3 – The system of geocentric directionals in Hiw

Out of seventeen languages spoken in the area, we will see that nine different systems of space directionals are attested – that is, nine different patterns of correspondences between geocentric and non-geocentric functions. While I will describe their similarities, I will also highlight their differences, and provide them with an interpretation. Of these nine systems, some – like the one in Dorig – are relatively straightforward and easy to explain; others – like Hiw – can only be unravelled based on a complex demonstration. Among the questions that need to be answered is the origin of *ag* in Hiw, the only directional whose meaning is, at least synchronically, purely geocentric. And then, how can one explain that this directional *ag* encodes both ‘landwards, inland’ and ‘towards SE’ – two directions which almost never align? Also, while Hiw and Dorig partly use the same directionals (‘up’–‘down’) in the hilly areas of their islands, they use completely different coordinates on the coastal zones. How can these differences be explained? Knowing that these space systems are historically related, how can their development be traced back? Among other things, I will show that the system of Dorig is conservative of Proto Oceanic, whereas the intricate paradigm of Hiw directionals results from several layers of historical innovations – some of which are shared with neighbouring languages, while others are specific to Hiw.

The present article will unfold as follows. Section 2 will present the Torres and Banks Islands, and introduce their nine systems of space directionals. Section 3 will describe their use of the northwest-southeast cardinal axis both across islands and within a single island—a domain for which all languages essentially work the same. Sections 4 to 8 will provide a systematic description of the nine geocentric systems found in northern Vanuatu, and analyse how they historically developed so as to become so diverse today. Finally, section 9 will synthesise these observations from the point of view of shared innovations and linguistic geography.

## **2 Paradigms of directionals in Torres and Banks Islands**

### **2.1 Northern Vanuatu landscapes**

The Torres and Banks Islands are two small archipelagoes located in the northernmost part of Vanuatu, with a total land surface of 882 km<sup>2</sup> (see the map in §2.2). Most islands are of volcanic origin,<sup>2</sup> with active volcanoes on Vanua Lava and Gaua. Their steep relief, covered in thick bush, rises up to relatively high cone-shaped summits, both in the Banks islands – Vanua Lava (921 m), Merelava (833), Gaua (767), Urepapara (764), Mota (411), Motalava (243) – and in the Torres islands – Hiw (366), Tegua (300), Toga (240), Lo (115). Several of these mountainous islands are surrounded by a more or less broad band of coral reef. Due to a geological process of uplift (Ballu *et al.* 2011), some islands even include accreted coral as their terrain for a fair portion of their surface. The flat, horizontal shape of those coral-based areas contrasts with the steep slopes of the central mountains.

The 9400 inhabitants of the Torres-Banks islands (VNSO 2009) are distributed across twelve of these islands, and approximately fifty villages. Some of these villages, especially on the higher islands, are located inland, on the slopes of the mountains where the soil is most fertile. But the majority of the modern population reserves the uphill areas for their subsistence gardens, and dwells in coastal areas – a convenient location where marine and land resources can easily be combined. The last two centuries have seen a trend for the population to leave inland hamlets, and settle down in larger coastal villages (Vienne 1984:23; François 2012:96-99).

In sum, the typical geographical setting of the Torres-Banks archipelago features steep mountainous cones covered in thick forests, sometimes surrounded by a flat coral terrain where coastal villages are found. This landscape can be represented in a stylised fashion, similar to the one given in Figures 2 and 3 above. Throughout this study, this diagram will constitute a useful background for the comparison of geocentric systems.<sup>3</sup>

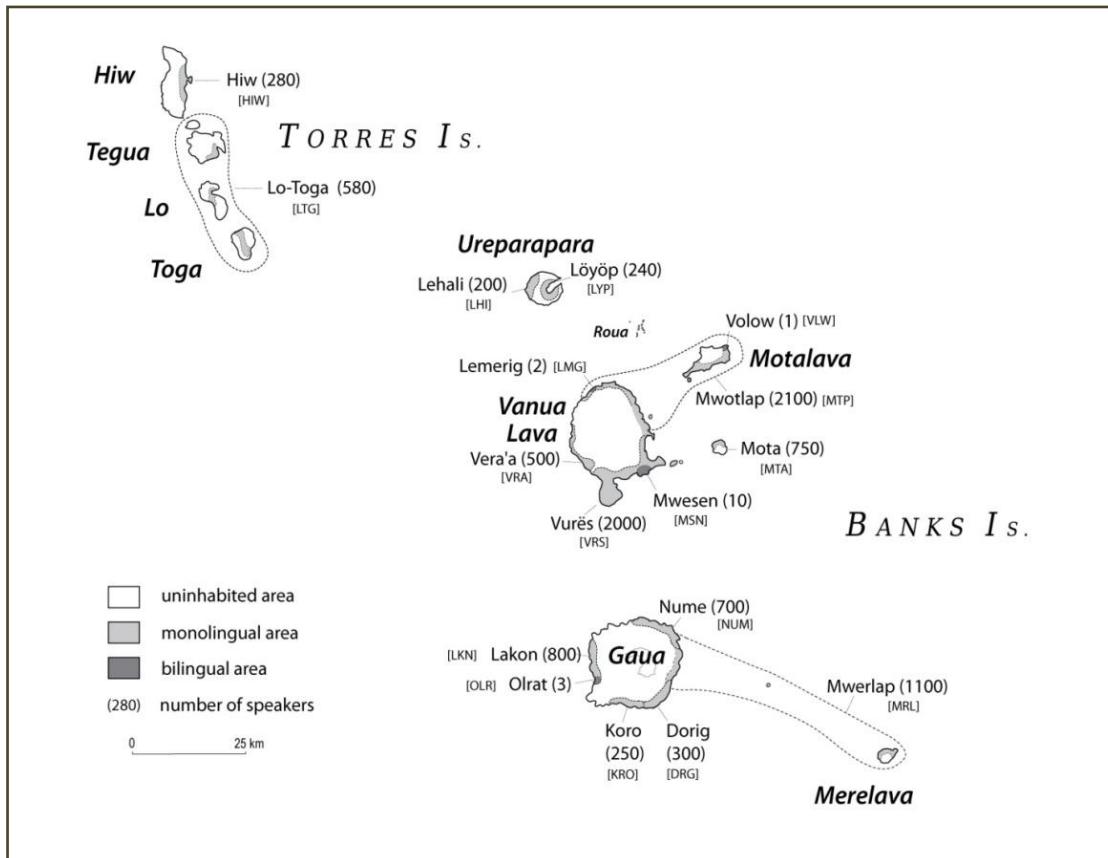
---

<sup>2</sup> The group of low islands known as Roua (or Reef islands), the only atolls of the area, were abandoned by its population in the middle of the 20<sup>th</sup> century (François 2012:97). The absence of any modern population there – apart from occasional fishermen from the neighbouring islands – makes it difficult to study the way geocentric directionals would be (or used to be) employed on these atolls, despite the obvious interest of such questions (Palmer 2007).

<sup>3</sup> I will discuss these diagrams again in §2.4.3 and §3.4.4.

## 2.2 The languages of the Torres and Banks Islands

Even though the Torres-Banks communities form together a relatively coherent social network, they speak 17 different languages. The present study rests on firsthand data which I collected on these languages, during a number of field trips in northern Vanuatu (since 1997 for the languages Mwotlap, Vurës, Mwesen; since 2003 for other languages). Map 1 provides a map of the area, showing the territory covered by the various language communities; clearly, coastal villages are the dominant form of settlement. Each language is given an approximate number of speakers, as well as a three-letter abbreviation.



Map 1 – The 17 languages of northern Vanuatu (Torres and Banks Is.)

These 17 languages are all Oceanic (Austronesian) languages, and together form the Torres-Banks “linkage” – i.e. they descend from what was once a dialect continuum (cf. Ross 1988; Kalyan & François f/c). During the three millennia of their *in situ* development, the communalects have diverged so much as to lose mutual intelligibility. However, these processes of diversification have always gone along a tradition of egalitarian multilingualism and social contact (François 2012), in ways which favoured various forms of cultural and linguistic diffusion. In this study, we will encounter such processes whereby innovations in the structure of space systems diffused across languages—e.g. the generalisation of the cardinal axis to all scales, which spread among all the ten northernmost languages of the area (§4.4). These issues of diffusion will also be the object of the final section (§9.2 – see Map 10 p.57).

### 2.3 A special paradigm of Space directionals

In a given language, spatial relations may contribute to the semantic organisation of more than one lexical category. For example, a set of adpositions may encode different topological relations (Eng. *in* vs *on* vs *under*); verbs can show lexical contrast based on deictic criteria (*go* vs *come*; *take* vs *bring*); a set of demonstratives may be based on the relative distance with respect to that centre, and so on and so forth. The same could be said of the Torres-Banks languages under study: spatial semantics are involved in various parts of their systems, and impact several of their word classes. The present study will focus on one particular paradigm of adverb-like particles called DIRECTIONALS.

Oceanic languages vary as to whether they express directional meanings using motion verbs (e.g. ‘go up’, ‘go down’) or directional adverbs (‘up, down’). Ross (2004, 2007:269) suggests Proto Oceanic may have had lexemes with dual membership, e.g. \**sipo* was both a verb ‘go down’ and a directional ‘down’. The languages of northern Vanuatu distinguish lexically between three word classes: they have directional verbs (e.g. Mwotlap *hēw* ‘descend, go down’); directional postverbs (Mwotlap *tēqēl* ‘downwards’); and a separate word class of directional particles (Mwotlap *hōw* ‘down’). While the three categories can perfectly combine in the same clause (e.g. *hēw tēqēl hōw* ‘go down’), the default strategy is to encode vectors using just the directional particle (e.g. *van hōw* ‘go down’). The present study will occasionally mention directional postverbs (§2.4.2, 4.2.2), but its main focus will be the systems of directional particles – or “directionals” for short – as they have properties of their own.

Grammatically, directionals form a subset of the larger class of *locatives*. In terms of glossing, these uninflected, usually monosyllabic particles are appropriately rendered by English adverbs such as ‘up’, ‘down’, ‘in’, ‘out’ – even though their semantics, of course, differ substantially from their English counterparts. Just like other locatives in these languages, the various syntactic functions of Torres-Banks directionals include that of locative predicate (cf. Eng. *She's down in the cellar*), of verb modifier (cf. Eng. *he walked down to the lake*), of noun attribute (cf. Eng. *the people down there*). While the morphosyntactic profile of directionals is generally similar in all languages in the region, there are also some language-specific peculiarities which will be described here when relevant.<sup>4</sup>

The primary function of directionals is to construct a spatial path or vector (François 2003:409). Sometimes, this vector reflects the direction followed by (a participant in) the reported event itself – as is the case in (3), from the language Lehali:<sup>5</sup>

(LHI.3)	Koyo	m-kal	<b>ila</b>	<b>ma</b>	l-eñ	ti	m-en	<b>hōw</b>	ti	m-mutuy.
	3du	PFT-cross	in	hither	in-house	then	PFT-lie	down	then	PFT-sleep
‘They came ( <b>in</b> ) into the house, lay <b>down</b> , and fell asleep.’									[LHI.Stepmother.63]	

In (3), the direction of the first motion event *kal* ‘cross [threshold]’ is encoded as ‘inwards’ (vector defined in topological, non-deictic terms) as well as ‘hither’ (vector defined in deictic terms). The second event *en* ‘lie’ is vectorised as a ‘downward’ movement.

<sup>4</sup> Section 6.3.2, for example, will describe the intricate morphosyntax of directionals in Mwerlap.

<sup>5</sup> Throughout this study, I will indicate the source of my examples using simple conventions. Sentences taken from recorded texts will note the language, the story and the sentence number – e.g. [Hiw.Meravtit.051]. Sentences obtained through elicitation refer to my field questionnaires – e.g. [LHI.d12:12]. Spontaneous speech heard during language immersion has a reference to my notebooks – e.g. [FP3-28b].

These same directionals are also used to refer to static locations, in which case their role is to define a vector between the deictic centre and that target location:

- (LHI.4) Qösö lē-vno **how** e tev lavēt.  
 HUM:PL in-village **down** DX IPFV celebrate  
 'People in the village **down** there are celebrating.'

[LHI.d12:12]

In (4), the directional *how* 'down' does not encode the direction of a motion event, but the orientation of the abstract vector that leads from the deictic centre ('here') to the location referred to. In sum, whether directionals encode a motion path as in (3), or serve to identify a static location as in (4), their principal function is to delineate a vector in a three-dimensional space.

## 2.4 Three strategies for directionals

After this definition of space directionals in grammatical and functional terms, I now describe the semantic contrasts around which these paradigms are structured. Directional systems in northern Vanuatu involve three types of coordinates: participant-oriented vs. topological vs. geocentric coordinates. I will review them successively in this section, before zooming into the geocentric type.

### 2.4.1 Participant-oriented directionals

All 17 languages have a pair of directionals which can be conveniently glossed 'hither' and 'thither'. The forms are given in Table 3, with languages ranked geographically from northwest (Hiw) to southeast (Mwerlap).<sup>6</sup>

Table 3 – Pairs of two participant-oriented directionals in Torres-Banks languages

	HIW	LTG	LHI	LYP	VLW	MTP	LMG	VRA	VRS	MSN	MTA	NUM	DRG	KRO	OLR	LKN	MRL
'hither'	me	me	ma	me	me	me	me	ma	me	ma	mē						
'thither'	vēn	vēn	van	van	va	van	wēl	suwō	net	nat	at	at	āt	āt	at	at	ot

#### ❖ EGOCENTRIC VS. ALLOCENTRIC

Beyond the convenient shortcut of their simple glosses 'hither' and 'thither', the definition of these two directionals normally includes the reference to a *participant*, typically animate, which provides the target of the spatial vector. Ross (2007:269) reconstructs Proto Oceanic as a system with four distinct deictic verbs: 'come, towards speaker' (\*mai); 'go towards addressee' (\*watu); 'go to; away from speaker' (\*lako), and 'go away' (\*pano). Even though some of these etyma are indeed reflected in the modern languages (§8.2, §11.3.2), the 17 languages of northern Vanuatu have reduced the system to a binary contrast between what I call an *egocentric* and an *allocentric* directional:

- gloss 'hither' = 'towards speaker'  
 an *egocentric* direction, targeted towards the speaker, or to any participant to which the speaker morally associates him- or herself.

<sup>6</sup> For a note on the etymology of these forms, see Appendix 2. I thank Stefan Schnell (pers. com.) for confirming the Vera'a form for 'thither'.

- gloss ‘thither’ = ‘towards non-speaker’  
an *allocentric* direction, targeted towards any participant that does not belong to the speaker’s sphere.

The allocentric directional typically encodes the directionality of a social event (e.g. transfer events like *give*, *show*, *speak*), as targeting either the addressee or a third-person participant; it usually translates in English as a dative (‘to you’ or ‘to him/her/them’). The word entails the instruction to retrieve a specific participant from the context, often without making it more explicit in the clause. Here are examples from Dorig, Mwesen:

- (DRG.5) Vu! Na s-aqtē **āt** i tsi-k nēk s-gān ni?  
god 1sg IRR-throw thither PERS same.sex.sibling-1sg 2sg IRR-eat 3sg  
'Spirit! Shall I throw (you) my sister, so you can eat her?'  
[lit. Shall I throw my sister *thither*...?]
- [DRG.Daughters.30]

- (MSN.6) Na ga van **nat**, piriñ kimi.  
1sg FUT go thither help 2pl  
'I'll come (to you<sub>PL</sub>) and give you<sub>PL</sub> a hand.'  
[lit. I'll go *thither* and help you<sub>PL</sub>.]
- [MSN.Qet.147]

Alternatively, the allocentric directional can be followed by a dative phrase explicitly indexing the referent – as in these examples from Lakon and Lo-Toga:

- (LKN.7) Apō ēhē lē uṁä, nēk ēt van **at** uh nē.  
Mum EXIST LOC house 2sg SUGG go thither for 3sg  
'Mum is in the house, you should go [*thither*] to her.'
- [BP5-15b]
- (LTG.8) Nie ni=ole ne=hih **vēn** hivie, pa nie ni=rōwetē.  
3sg AO:3sg=take ART=breast thither DAT:3sg but 3sg AO:3sg=reject  
'She tried to give him [*thither*] some milk, but he wouldn't have it.'
- [LTG.d02:23]

Even though the target of the allocentric directional is usually animate, it can be an inanimate referent that is salient in the context (example in Volow):

- (VLW.9) Din noyon e Welan. Nēñ tivele ġalġalbat **va** ben!  
room POSS PERS Chief 2sg PROH RED~enter thither in.it  
'This is the Chief's room. Don't ever go in there!'
- [VLW.Orphan.85]

This possibility for the directional to target inanimate participants is the reason why I choose to label this type of coordinates “participant-oriented” – as opposed to “personal” as in François (2003).

Like the allocentric directional (‘thither’), the egocentric one (‘hither’) is typically associated with social events when they are directed towards the speaker, or towards the main character of a story, etc. Here are examples in Hiw and Lemerig:

- (HIW.10) Ārōñ te **me** ne!  
hear SUGG hither PROX  
'Listen (to me)!' [lit. Listen *hither*!]
- [Hiw.d12:01]
- (LMG.11) Ē Vārvañ e ti n-row **me** sur ē Qet e n-lāñ ti.  
PERS (name) DEF 3sg AO-jump hither DAT PERS (hero) TOP AO-strike 3sg  
'The Giant jumped [*hither*] at Kpwet and whacked him.'
- [LMG.Qet.214]

### ❖ COMPETING STRATEGIES AMONG DIRECTIONALS

An important property of these participant-oriented directionals is that they come in complementary distribution with other types of directionals.

A detailed study of the Mwotlap system (François 2003) showed that speakers, when encoding vectors using directionals, have to choose between the three types of directionals outlined above, each corresponding to a different strategy: the “personal” strategy; the “topological” strategy; and the “geocentric” strategy. This choice follows a hierarchy, whereby the personal (or *participant-oriented*) strategy will always take priority on the other directionals: that is, even if any vector could potentially be encoded in purely spatial terms, there is a consistent preference for anchoring the vector onto a participant, whenever this is contextually available. As a corollary, the participant-oriented directional ‘thither’ – like *van* in (12) – never combines with spatial directionals, but replaces them.

- (MTP.12) Hayveg ***van***.  
 enter thither  
 PARTICIPANT-ORIENTED STRATEGY: ‘Go in! (towards him/her)’ [implies someone is inside]
- (MTP.12') Hayveg ***hay***.  
 enter in  
 TOPOLOGICAL STRATEGY: ‘Go in!’ [suggests nobody is inside]

### ❖ THE TWO USES OF ‘HITHER’

Finally, an extra complication is the fact that the directional ‘hither’ is used in two different ways. One meaning of ‘hither’, the one illustrated in (10) above, is to point to EGO (or another participant of the EGO sphere) as the target of a participant-oriented social event. In this sense, ‘hither’ contrasts with ‘thither’, and it replaces space directionals rather than combine with them: a combination such as the one in (10’) would be impossible in Hiw.

- (HIW.10') ?? Řōn̄ te iy ***me*** ne!  
 hear SUGG in hither PROX  
 ?? ‘Listen! (to me, who am standing inside)’

However, the same directional ‘hither’ is also used in northern Vanuatu languages with a meaning ‘towards the deictic centre’. In this sense, the vector is not construed with respect to a participant in a social event, but is a purely spatial vector defined with respect to ‘HERE’ as a spatial point. This non-personal interpretation is the only case when ‘hither’ can follow another directional – as in (3) above. Unlike (10), the directional *me* in (13) does not refer to EGO as the target of a participant-oriented event (as in ‘who visited me?’); rather, it encodes a pure spatial vector towards HERE, the location of the speaker at the moment of utterance. A construction like (13) is well adapted to a situation where the house was empty at the moment of the visit:

- (HIW.13) Ye ve tō ti ***iy*** ***me*** yōñwe?  
 who PFT<sub>1</sub> go PFT<sub>2</sub> in hither in.house  
 ‘Who came in this house?’

[Hiw.Grouper.33]

This is the only case when two directional particles can combine. Such a possibility is restricted to ‘hither’, and not open to ‘thither’:

(HIW.13')	Ye ve tō ti <b>iy</b> (*vēn) yōñwe?
	who PFT <sub>1</sub> go PFT <sub>2</sub> in (*thither) in.house
	'Who went in the house?'

In sum, there are really two functions of 'hither'. First, it is a participant-oriented directional pointing to EGO (or assimilated), in which case it contrasts with 'thither' – cf. (10). Second, it is a spatial deictic pointing to HERE (or assimilated), in which case it contrasts with the use of a spatial directional alone – cf. (13)-(13'). This semantic distinction between the two functions of 'hither' is a covert one in the languages of northern Vanuatu, and correlates with no formal difference in the directional itself; I will therefore use the same gloss 'hither' in both cases. The different functions of participant-oriented directionals are summarised in Table 4; in this table, "space directional" refers to any directional (topological, geocentric) that does not belong to the pair of participant-oriented directionals.

*Table 4 – The functions of participant-oriented directionals*

	<i>Participant-oriented vectors</i>	<i>Spatial vectors</i>
<i>Egocentric direction</i>	HITHER alone – cf. (10)	(space directional) + HITHER – cf. (13)
<i>Allocentric direction</i>	THITHER alone – cf. (5)-(9)	space directional alone – cf. (12'), (13')

Because the present paper is focused on the relations between topological and geocentric directionals, the participant-oriented directionals won't be discussed any further here.<sup>7</sup> They were mentioned here because they form part of the same closed set of directionals, and contrast paradigmatically with the geocentric ones.

#### 2.4.2 Topological directionals

I now turn to what I call "topological" directionals, a category which includes two pairs of directionals. The pair *in–out* is defined by reference to a closed shape interpreted as a container: house, canoe, basket, pocket, etc. The second pair *up–down* is defined with reference to the vertical axis.

In the terminology used by Levinson (1996b) and illustrated in Figure 1 above, the term "topological" only refers to the first of these pairs (*in–out*), whereas the vertical dimension is treated separately. However, as we will see, the languages of northern Vanuatu treat the two pairs of directionals as members of a single subparadigm, and it is therefore legitimate to group them under a single category, for which the label 'topological' is well adapted.<sup>8</sup>

Table 5 provides the forms of the topological directionals for the 17 Torres-Banks languages. The reader will find in the Appendix comments on the pronunciation of these forms (§11.2), as well as notes on their etymology (§11.3).

<sup>7</sup> I will however mention 'thither' again in the section on Hiw (§8.2), which colexifies this meaning with 'southeast on land'.

<sup>8</sup> Levinson himself (1996b:360) acknowledges that "the VERTICAL dimension is special in various ways and is an angular specification that creeps into essentially nonangular TOPOLOGICAL specifications" (my emphasis).

Table 5 — Topological directionals in Torres-Banks languages

	HIW	LTG	LHI	LYP	VLW	MTP	LMG	VRA	VRS	MSN	MTA	NUM	DRG	KRO	OLR	LKN	MRL	
'in'	iy	il	ila	say		ha	hay	sar	sar	sar	sar		sage	sa	sag	sa	saa	
'up'	vēn	vin	vēn	sa			hag	sag	sag	siag	sag			roka		hag/	sar	
'down'	uw	iw	how	sōw	hō	hōw	sōw	suwō	sōw	sōw	swo					hōw/	sōw	
'out'	rōw	rōw	yow	yow	yo	yow	row	rōw	rōw	row	rowo		ror	ror	ror	roy	rōkōw	row

Most of the systems have four topological directionals. Mwesen is such a language:

- (MSN.14) E Qet ni le o gepen no, mop kal **sag** le ak.  
 PERS (hero) AO take ART sail DEF put upwards **up** LOC canoe  
 'Kpwet took the sail and put it **up** on the canoe.' [MSN.Qet.031]
- (MSN.15) Me rovrov o parpar, qēs **sōw** le qiti ak no.  
 PFT IPFV~raise ART axe smash **down** LOC head canoe DEF  
 'He raised his axe, and smashed it **down** onto the canoe's prow.' [MSN.Qet.085]
- (MSN.16) Kal telñor **sar** le gemel, nē ni on le tenepa-n.  
 cross inwards **in** LOC dwelling 3sg AO lie LOC bed-3sg  
 'He walked **into** his dwelling, and lay down on his bed.' [MSN.Varvang.50]
- (MSN.17) Ni o-o-on le lo gemel, ni row lō **row** le sar.  
 AO DUR-lie LOC inside dwelling AO jump outwards **out** LOC clearing  
 'He remained in his home for a while,  
 but suddenly rushed **out** to the frontyard.' [MSN.Varvang.47]

Note, in passing, that directionals are commonly preceded by a word with a similar meaning: *KAL sag* in (14); *TELÑOR sar* in (16); *LŌ row* in (17). These terms, briefly mentioned in §2.3, belong to a separate word class, that of verb-modifying *postverbs*. They have different distributional properties; and their function is only to specify the path of a motion event, never to define a static location – contrary to directionals which can have both functions. While these postverbs, in principle, do not fall under the scope of this study, it may be useful to underline their optional presence. These verb modifiers are not attached to any spatial strategy in particular: for example, the one glossed 'upwards' (*kal* in Mwesen) can be used with a vertical meaning, but also with the other functions attached to 'up', such as the cardinal 'up' pointing to southeast (§3.3). When postverbs and directionals are used in the same clause, the normal situation is for them to semantically align: 'outwards' with 'out', 'upwards' with 'up'... with only a few exceptions (François 2003:426). I will mention these postverbs again in §4.2.2.

As Table 5 shows, some languages have fewer than four topological directionals. In the case of Volow, a language variety very close to Mwotlap, the collapse between 'in' and 'up' seems to be due to a more general pattern of morphological truncation which has affected all the directionals (\**hōw* → *hō*; \**yow* → *yo*; \**van* → *va*) and which has resulted in the loss of distinction between 'in' (\**hay* → *ha*) and 'up' (\**hag* → *ha*).

This accident of historical morphology in Volow contrasts with the situation in the five

languages of Gaua, for which the colexification<sup>9</sup> is systematic, on the one hand, between ‘in’ and ‘up’; and on the other hand, between ‘out’ and ‘down’. Only the context makes it clear which of the two coordinate axes is being meant in a particular utterance. To take an example from Olrat, the directional *saa* corresponds to a vertical ‘upwards’ movement in (18) (cf. ‘lift’, ‘lintel’); whereas in (19), it means ‘(look) inwards’:

- (OLR.18) Ni mō sēj rakat ni **saa** lē mataalol.  
 3sg PFT hang lift 3sg **up/in** LOC lintel  
 ‘He hanged (the ogress) **up** above her doorway.’ [OLR.Ogress.082]
- (OLR.19) Nōrō tē pipira ti nōtam, nōrō mō pipleñ **saa** lē vuvuy.  
 3du IPFV<sub>1</sub> play IPFV<sub>2</sub> outside 3du PFT peer **up/in** LOC house  
 ‘As the two boys were playing outside, they looked **into** the house.’ [OLR.Eel.36]

In a similar way, the reverse polysemy characterises the Lakon directional *hōw* ‘down/out’: in (20) it is used vertically to mean ‘down’, but in (21) it refers to an outward movement:

- (LKN.20) Ni tē tārā rāgā neñ tē sēv **hōw** lē tanē.  
 3sg SEQ chop tree DEF SEQ fall **down/out** LOC ground  
 ‘He chopped the tree, which fell **down** on the ground.’ [LKN.d07:03]
- (LKN.21) Ni 'n sapläg tōpō-n ni tē rowol **hōw** matumā.  
 3sg PFT carry.child grandchild-3sg 3sg SEQ cross.door **down/out** frontyard  
 ‘She took her grandchild in her arms and walked **out** to her frontyard.’ [LKN.d02:28]

This pattern of colexification is characteristic not only of the five languages of Gaua, but also of a number of other languages in the Oceanic family. For example, Ozanne-Rivierre (1997:86; 1999:79) reports that in Oceanic languages of New Caledonia, “the up/down oriented axis is used to express (...) *inside the house* vs *outside the house* and, when one is *inside the house*, *towards the interior of the house* vs *towards the door*.“ She goes on to suggest a link between this pattern and the traditional architecture whereby houses used to be built on stilts.

Historically, it is quite possible that these colexifications ‘up/in’ and ‘down/out’ were characteristic of Proto Oceanic itself. Indeed, Ross (2007) does not reconstruct any form for ‘(go) in’ or ‘(go) out’, and states “It is reasonably clear that the ‘inside’/‘outside’ opposition found in European languages did not occur in POc” (2007:255). By contrast, POc clearly had terms for ‘(go) up’ (\**sake*) and ‘(go) down’ (\**sipo*) – see §3.2 below. Besides, the etyma \**saro* ‘in’ and \**rowo* ‘out’ which can be reconstructed for the Torres-Banks area (see the Appendix in §11.3) do not seem to be attested anywhere else in Oceanic languages. All this suggests that the languages of Gaua, with their two-term system ‘up/in’ vs ‘down/out’, may in fact be conservative of the system of Proto Oceanic. Conversely, the existence of four-term systems with separate lexification of ‘in’ and ‘out’ reflect most probably a local innovation, which diffused across the whole Torres and Banks area, yet left Gaua untouched.

Finally, Mota presents a hybrid and unusual situation. It shares with other Banks languages the innovative directional *rowo* ‘out’, yet it behaves like Gaua in lacking a specific directional for ‘in’, which it still colexifies with ‘up’:

---

<sup>9</sup> On the concept of *colexification* – i.e. pattern whereby a language lexifies two senses with the same form – see François (2008).

(MTA.22)	Tamate ilon ni me sarovag pata	<b>sage</b>	lele iñma.
	ghost DEF 3sg PFT enter inwards up/in		inside house
'The ghost came into the house.'			[MTA.GhostSister.28]

In §4.3 and 9.1 below, I will show that these innovations of the topological system had important consequences on the encoding of geocentric directions.

### 2.4.3 Geocentric directionals

So far, we have seen how the directionals of northern Vanuatu can encode vectors based on a particular participant ('hither' vs 'thither'), or on the topological configuration of a particular setting, using either the vertical axis or an 'in'—'out' contrast. While these first two strategies are unproblematic, they are unable to encode one very common type of spatial vector – that is, those vectors belonging to the horizontal plane which are defined neither by their relation to a participant, nor to their position in a predefined topological configuration.

In order to encode spatial directions on the horizontal plane, northern Vanuatu languages resort neither to intrinsic nor to relative strategies (see §1.1), but to an absolute frame of reference defined by geocentric coordinates. But such systems cannot be reduced to a simple four-way symmetrical system such as *North-East-South-West*: instead, they involve a number of different axes and subsystems whose detail differs across languages. In a nutshell, the various directions which northern Vanuatu languages encode involve the following axes:

- a fixed cardinal axis oriented southeast–northwest
- a land-sea axis running orthogonal to the shore, employed on land or at sea
- an axis running parallel to the shore, whose general orientation ( $\pm 90^\circ$ ) is either SE or NW
- an axis oriented *uphill* vs *downhill*, and used in the higher areas of certain islands

On top of these geometric distinctions, some systems introduce lexical oppositions based on scale: for example, some have different directionals for 'towards SE' depending on distance.

Crucially, these different directions are not all encoded with separate forms: instead, each language colexifies some of these meanings together. For example, we saw in §1.2 that Dorig [Table 1] uses the same directional *sag* for 'in', 'inland' and 'parallel to shore, towards SE'; conversely, Hiw [Table 2] lexifies differently 'in' (*iÿ*) from 'inland' (*ag*), but conflates lexically 'parallel to shore towards SE' with 'thither' (*vĕn*). In order to make these complex systems cross-linguistically comparable, I will follow the typological approach to polysemy exposed in François (2008), and choose to represent each potential vector as an atomic sense in an "etic grid"; this will enable us to observe how these senses are being colexified by each language.

Table 6 provides an overview of all directional systems in Torres and Banks Islands. All three domains of use are mentioned here. The two top rows reproduce the participant-oriented directionals we saw above (see Table 3). The rows with grayed headings (#3, 7, 14, 18) correspond to the topological directions that were discussed in §2.4.2. All other rows correspond to the various types of geocentric directionals which are used in the absolute strategies of space reference. The different rows are organised in such a manner that all the senses which are colexified in a given language should be adjacent in the table.<sup>10</sup>

---

<sup>10</sup> The only exceptions to this principle are: the row for 'thither', which shows certain polysemies in some languages (see §8.2); and the Hiw form *ag*, which is indeed idiosyncratic in its meaning (§8.3).

*Table 6 - Systems of space directionals of Torres and Banks languages: a comprehensive chart  
(languages ranked geographically from northwest to southeast)*

For future reference, Figure 4 shows the correspondences between the different rows of Table 6 (number codes, marked with '#') and the vectors used on the graphic representation of directional systems (see Figures 2-3 p.5). Its objective is to facilitate the reading of Table 6, and associate each vector with the corresponding forms. For example, vector #5, "*inland (as used typically in the context of a coastal village)*", is lexified as shown in row #5 of Table 6: **ag** in Hiw, **il** in Lo-Toga, **ila** or **la** in Lehali, **say** in Löyöp, etc.

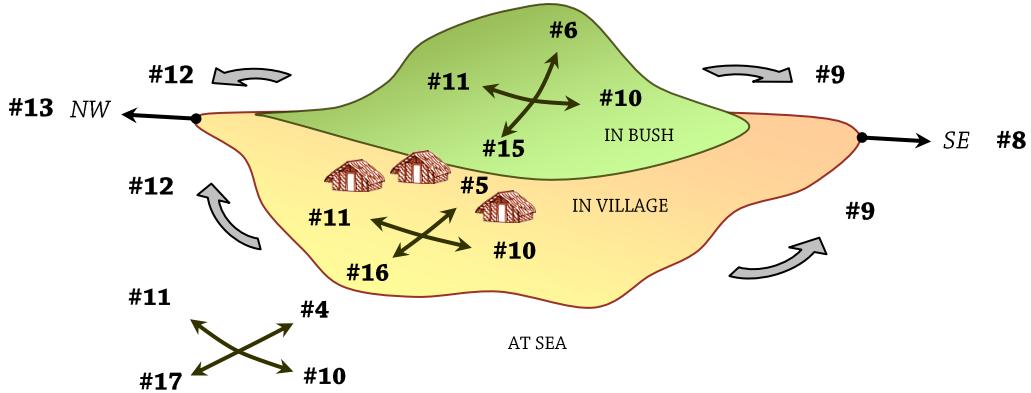


Figure 4 – Number codes for the various vectors under discussion, as listed in Table 6.

#### 2.4.4 Unity and diversity

The various systems displayed in Table 6 share a family resemblance, and one may be sensitive to the amount of characteristics that are shared by all these languages. For example, all languages colexify #14 *down* with #12 *towards NW (for long distances)*; all languages colexify #18 *out* with #16 *seawards*; and so on. This being said, the key point of the present study is really the amount of diversity found in the space systems of the region. For such a small set of languages, this variety is surprising – especially given the amount of contact-induced structural isomorphism that is typically shown by these languages (François 2011). Obviously, the diversity of these systems meant that I had to observe each language in its own terms, with no aprioristic view of which semantic contrasts or space strategies should be expected.

It is however not the case that these 17 languages define 17 different systems. The four languages of Vanua Lava (Lemerig, Vera'a, Vurës, Mwesen), for example, have the same geocentric directionals, with just a few phonetic differences. The same can be said of the five languages of Gaua: whether their directionals are historically cognate or not, they operate the same structural contrasts. Finally, Lo-Toga and Lehali also show isomorphic space systems, even though they are spoken on different islands.

If one decides to distinguish directional systems on the basis of their structural contrasts rather than on the actual forms of their words, then one must conclude that the 17 languages of the Torres-Banks islands define a total of **nine** different systems of geocentric space reference. It is the purpose of the main sections of this study (§3-8) to examine them one after the other, and unravel the history of their development.

### 3 The navigational scale and the NW-SE cardinal axis

#### 3.1 Local scale vs navigational scale

In accordance with earlier studies of similar Oceanic systems (Ozanne-Rivierre 1997; Hyslop 2002; Palmer 2002:128; François 2004; Ross 2007:229), it is necessary to draw a preliminary distinction between two scales of geocentric reference, as they typically involve distinct directional subsystems.

On the one hand, the *local* scale corresponds to those vectors, locations and directions, that belong within a radius of about 200 meters around the centre of reference – typically in the setting of the village. On the other hand, the *navigational* scale corresponds to long-distance vectors, prototypically those defined at sea, or across islands.

This section 3 focuses on the navigational scale, and in particular the use of the cardinal axis oriented NW-SE. As for the intermediate scale – the one that involves, for example, the distances between two villages on the same island – it essentially belongs to the navigational domain, yet its anchoring on land entails some specific characteristics distinct from its use across islands. Section §3.4 will precisely examine how languages adapt the cardinal axis to the shape of islands.

#### 3.2 The navigational scale in Oceanic languages

Based on earlier scholarly work, François (2004) proposed a systematic comparison of sixteen Austronesian languages and their space systems, and outlined their commonalities and differences; these will be summarised here. Oceanic languages show remarkable consistency regarding the navigational scale. Virtually every Oceanic language employs a single cardinal axis that is oriented NORTHWEST — SOUTHEAST. Some of the exceptions can be explained by local adaptations to the topography of a particular island – like Mwotlap which has rotated its axis to a west vs east direction, due to the topographical orientation of its main island Motalava – see (1)-(2) above, and the discussion in §3.4.3. However, even though Mwotlap has rotated its cardinal axis on its main island, it orientates it consistently NW-SE in all other contexts (François 2003:428–434). This confirms the present relevance of this cardinal axis even in those languages which seem to have changed it.

Everywhere, this single cardinal axis is lexified UP (southeast) vs DOWN (northwest), using the terms used for vertical coordinates. Most modern speakers are unable to explain the reason for such a pattern; some suggest a possible connection with the rising and setting of the sun. However, both the orientation of the axis (SE-NW rather than E-W) and the semantics of the *up-down* contrast argue in favour of an alternate analysis in terms of **winds** (Ozanne-Rivierre 1997:85; François 2004:11). In the terminology of ancient Oceanic navigators, the difficulty of sailing against the southeast trade winds was assimilated to travelling ‘upwards’, as opposed to the easy ‘downward’ navigation that was done towards northwest, following the wind. In other terms, the vertical terms *up* and *down* in Oceanic languages were given the same semantic extension as the one found with English *upwind* and *downwind*. In the remainder of this paper, I will occasionally refer to the terms *up* and *down* of this cardinal axis using precisely their English gloss *upwind* and *downwind* – even when modern speakers have forgotten the original connection with seafaring terminology.

Based on this observation of modern Oceanic languages, François (2004) proposed to reconstruct the space system of Proto Oceanic. As far as the navigational scale is concerned,

a single cardinal axis oriented NW-SE was lexified using the vertical terms *\*sipo* '(go) down' and *\*sake* '(go) up' – see Figure 5. In such a navigational system, the world is divided in two halves: with respect to any point in the Pacific, all islands located in the northwest half will be located 'down', and those located southeast will be 'up'. This pair was possibly supplemented by the verb *\*pano* 'move in a transverse direction' (Ross 2007:279) for directions that were neither 'up' nor 'down'.

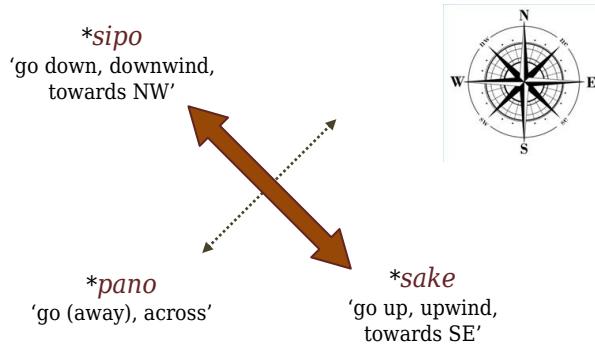


Figure 5 — The navigational scale of Proto Oceanic (after François 2004:20)

### 3.3 The cardinal axis in northern Vanuatu

The 17 Torres and Banks languages have preserved the cardinal axis of their ancestors. Everywhere, it takes the form of a single axis oriented NW-SE, and lexified using the vertical directionals 'down' and 'up'.

To be more exact, the pattern of colexification is attested everywhere for the *down* directional (i.e. rows 13 and 14 of Table 6 always match), and it is also found in most languages for *up* (see rows 7–8 of Table 6). The only exceptions are the first three languages, which show separate lexification for '(vertical) up' and 'towards SE'. I will argue that this modern situation is due to a lexical innovation for '(vertical) up', whereby a root *\*vene* replaced the old directional *\*sake* (cf. §7, 8.3). Even though the latter etymon has been lost for the vertical sense of 'up', it survives vestigially with its geocentric functions: thus Hiw *ag*, Lo-Toga (*i**ag*) and Lehali *ha*, which in synchrony must be glossed 'towards SE (on the navigational scale)', are all historical reflexes of POc *\*sake* '(go) up' (§11.3.4). In this sense, Torres-Banks languages are no exception to the observation that the cardinal axis 'down[wind]–up[wind]' is everywhere lexified using vertical directionals 'down' and 'up' – at least etymologically.

These languages also confirm the general NW-SE orientation that has been observed in most other parts of the Pacific. In order to limit the impact of a particular island's local topography, the best way to observe this cardinal axis is done by locating oneself on one island, and pointing towards other islands (see Map 1).

Someone located on the island of Lo, in the middle of the Torres group, will 'go down' to Tegua (LTG *vēn iw Tugue*) or Hiw (*vēn iw Hiw*) but 'go up' to Toga (*vēn iag Toge*); 'up' will also be the direction of all islands located southeast of the Torres, such as the Banks islands (*vēn iag Venie Gōo* 'go up to Ureparapara', *vēn iag Venie Tewale* 'go up to Matalava', etc.). Interestingly, this NW-SE orientation shows up in the toponymy: the name given in Lo-Toga to the Banks group is *Sag*, etymologically from POc *\*sake* 'up'; and the very name of the island *Hiw* historically meant 'down', from POc *\*sipo*.

Obviously, the limit between ‘up’ and ‘down’ will depend on the point of reference. Thus the island of Vanua Lava, which is located ‘up’ from the Torres Is, will be described as ‘down’ by someone located on Gaua (e.g. Lakon *van hōw Vanōlav* ‘go down to Vanua Lava’). From Gaua, the UP direction will point to all islands further south or southeast (e.g. Dorig *van sag Merlav* ‘go up to Merelava’, *van sag Vila* ‘go up to Vila’; Lakon *van hag Mew* ‘go up to Maewo’, *van hag Sajo* ‘go up to Espiritu Santo’), etc. The same demonstration could be made for Mwotlap (François 2003:432) or other languages of the area, since they all work the same way.<sup>11</sup> Figure 6, inspired by Hyslop (2002:49) for Ambae, illustrates the situation that obtains for a speaker of Löyöp on Ureparapara, with its pair of directionals *sa* ‘up’ vs *sōw* ‘down’.

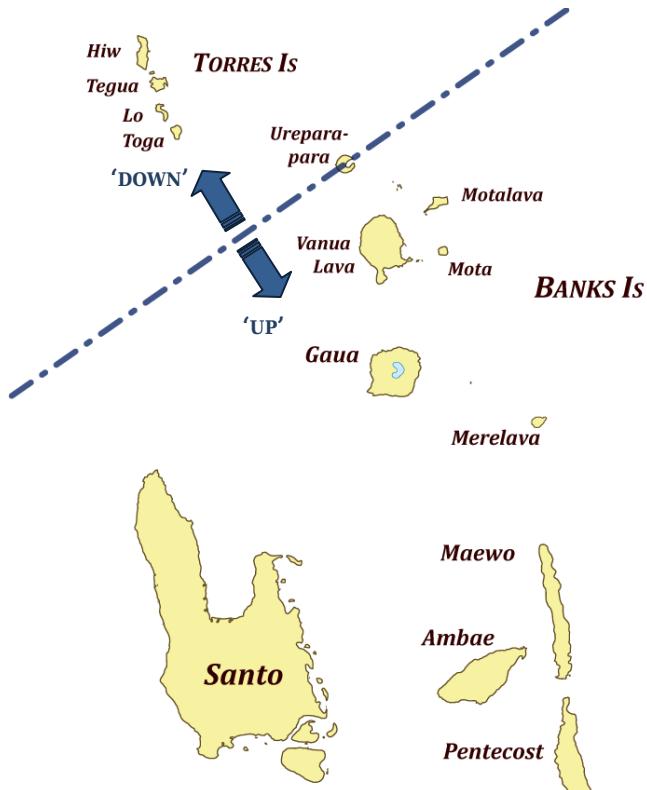


Figure 6 — The cardinal axis for referring across islands: ‘up’ and ‘down’ directions for a speaker located on the island of Ureparapara

Cardinal directions between islands are mentioned in everyday conversation, and also feature in numerous traditional narratives. For example, the famous mythological hero Qat or Kpwat (cf. François 2011:184; f/c) left his home on Vanua Lava, and sailed *down* to Vava – the ancient name of the Torres islands – where he bought the Night; then he sailed back *up* to his island. In another myth, Kpwat’s eleven brothers stole his canoe on Vanua Lava, and paddled it *up* to Gaua; from there, they had to find a way to go back *down* to their home island. The origin myth of the yam as it is told in the Torres Is tells how it grew in Hiw, and how its creeper vine went *up* to the Banks, all the way *up* to Pentecost and Malakula in the south – see ex.(23):

<sup>11</sup> The languages of Vanikoro, an island of the Solomons located just north of Hiw, use their cardinal axis in much the same way as Torres-Banks languages (François 2009:117). Thus in the main language Teanu, the term *Iura* – literally ‘upwards’ – designates the Torres-Banks islands and the whole of Vanuatu further south.

(LTG.23) Ne=keka in ni= vēn dē ne= Lō ni= vēn **iag** Toge;  
 ART=yam DEF AO:3sg= go from ART= Lo AO:3sg= go (up,SE) Toga  
 ni= vē **iag** Perepare... ni= vēn wahe Melakule.  
 AO:3sg= go (up,SE) Ureparapara AO:3sg= go until Malakula  
 'The yam (creeper) left Lo, went south(east) to Toga;  
 it went south(east) to Ureparapara...; and all the way to Malakula.' [LTG.Yam.07]

These narratives, combined with ordinary conversation, constitute the context in which knowledge of the *up-down* cardinal axis is transmitted from one generation to the next. Indeed, even though linguists can demonstrate that this axis is historically based on the main trade winds that were so significant to ancestral navigators, this association of directionals with winds has been lost today virtually everywhere in north Vanuatu. Speakers on Motalava grow up on their home island, in a social environment where the island of Gaua is always associated with the directional 'up' (**hag Alkon**), and Ureparapara always with 'down' (**hōw Nōybaybay**), following patterns of collocation which are lexicalised and repeated as such. Knowledge of the most frequent collocations allows them to abstract away a general orientation of the axis based on known directions, without ever having to refer to the direction of the wind, the sun or any other bearing other than actual islands and places.

Unlike the European N-W-S-E system which defines four quadrants, the *up-down* cardinal axis divides up the world in two halves, respectively the SE half ('upwind') and the NW half ('downwind'). It sometimes happens that the island one points to is located neither clearly southeast nor clearly northwest; in this exceptional case, languages resort to distinct directionals, typically those used for the land-sea axis. Thus for people of the low island of Motalava, the higher island of Vanua Lava to their southwest is neither 'up' nor 'down', it is 'landwards' (**hay Apnōlap**) - see François (2003:432).

Like the European cardinal directions, the cardinal axis is abstract and is not bounded at any particular point. On the 'down' side are the Solomon Islands; on the 'up' side are all the other islands to the south, including the towns of Santo and the capital Vila. That the capital is 'up' is purely due to geographical reasons: the hypothesis that one goes 'up' to the capital due to the supposed prestige of the city (as in Eng. *move up to London* or Fr. *monter à Paris*) would be ethnocentric, and at odds with what we otherwise know of these egalitarian, horizontal societies. If a Torres-Banks speaker were located on Tanna island and pointed to the capital to their north, they would refer to it as 'down'. Likewise in Teanu of the Solomon Islands (note 11), the capital Honiara is located 'down' (**tev' tawo**) from Vanikoro. This example from a neighbouring area suggests that the location of capital cities on the *up-down* axis has nothing to do with any symbolic value attached to *up-down*: it is merely indexed on the geography.

Western countries - France, Britain, Australia, Japan... - are also all located 'up' (Mwotlap *van hag Japan* 'travel up to Japan'), even when their actual location is northwest from Vanuatu. Again, one could interpret this as a token of respect that would be supposedly attached to vertical terms, as though pointing to foreign countries were somehow assimilated to looking 'up'. However, no evidence would back such a hypothesis; I believe the answer simply has to be sought in the geography as it is *perceived* by social actors. Indeed, the common experience of Torres-Banks people is that the only way to go to or come from these foreign countries involves a trip via the capital Port Vila, which is situated south: therefore, anyone leaving the area in order to go abroad will first have to head south, i.e. 'up'.

### 3.4 Adapting the cardinal axis on land

The use of the cardinal axis at sea, or across different islands, is straightforward: everywhere, the axis is oriented NW-SE. Things become more intricate on land. Sections 4 to 8 below will examine the different systems on the *local* scale, i.e. the directionals used for short distances (e.g. within a single village); as we'll see, some languages make use of the cardinal axis in that context, while others do not. But first I propose to discuss not the local scale, but the "intermediate" scale: namely, those directions that still involve long distances, yet take place within a single island. This scale will typically be used when pointing from one village to another on the same island; or when talking about one's journey walking across the island.

Long-distance navigation within one island is expressed essentially using two axes. The most salient axis on land is the *land-sea* axis, a variable direction that radiates from the centre of the island, in all directions, towards the sea. If I am standing in an inland hamlet and I refer to a village down on the coast, chances are I will be using the directional for 'seawards' – either *out* or *down*, depending on the language – and vice versa (I will come back to the *land-sea* axis in §4). But villages on one island tend to be typically located on the same altitude, whether along the coast or on a plateau. In this case, the type of vectors needed to encode the direction from one village to another will involve directions on the horizontal plane.

Starting from any location on the island, two axes can be defined: one defined by the declivity of the ground (*land-sea*); and a second axis for all directions PARALLEL TO THE SHORE. Torres-Banks languages express these long distance directions parallel to the shore using the cardinal axis UP-DOWN. The corresponding directionals can be seen in Table 6 (p.16), on rows #9 and #12. These can be compared, respectively, with rows #8 and #13: in almost all languages, the cardinal terms used for long distances on land are identical – not too surprisingly – to the ones used at sea.<sup>12</sup>

#### 3.4.1 Skewing of cardinal directions on land

In principle, the cardinal axis on land should have the same orientation, in compass terms, as the NW-SE axis that is used *across* different islands (§3.3). However, one major difference is that, on land, spatial orientation is preempted by the contrast between *land* and *sea*: due to its high perceptual salience (Palmer 2002:114), the *land-sea* axis is always the primary axis of the orientation system. As for the cardinal axis, it only receives a secondary status on land (François 2004:12-14): its orientation is always redefined so as to run orthogonal to the main *land-sea* axis. As a result, the cardinal directionals *up* and *down* are always used on land for directions that run parallel to the shore.

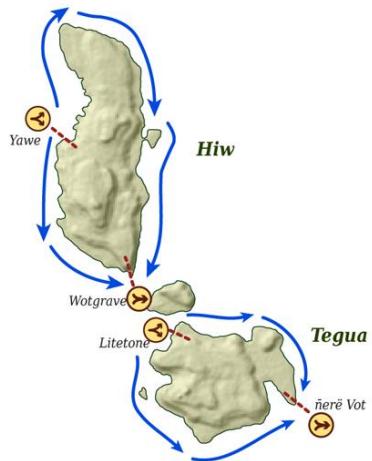
Of course, this orientation parallel to the shore entails that, even though the cardinal axis does sometimes align with proper coordinates NW-SE (with SE as 'up'), in many places its actual orientation, in compass terms, will be skewed by the shape of the shoreline. This is how the directional 'up' can point due South in some points, due East in others, or even ENE or SWS elsewhere.

---

<sup>12</sup> The only exception is Hiw, which has *ag* for cardinal SE across islands, but *vēn* (lit. 'thither') for cardinal SE on land (see also Figure 3 p.5). I will come back to this complex system in §8.

### 3.4.2 The paradox of cardinal convergence and divergence

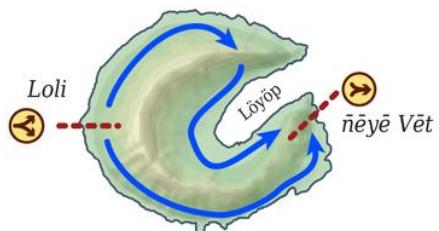
The following maps indicate the orientation of the cardinal axis UP-DOWN on the eight main inhabited islands of the Torres and Banks islands. All blue arrows point towards cardinal UP; cardinal DOWN is not indicated, as it simply corresponds to the reverse direction along the same arrows. The other symbols I use will be explained below.



Map 2 - UP[wind] directions  
in **northern Torres Is**



Map 3 - UP[wind] directions  
in **southern Torres Is**



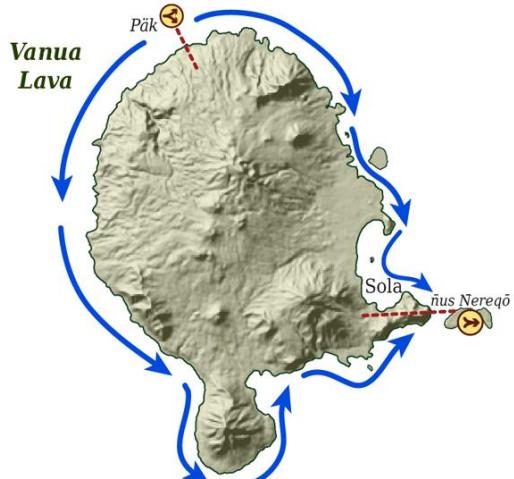
Map 4 - UP[wind] directions on **Ureparapara**



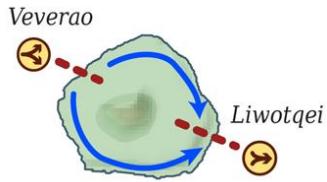
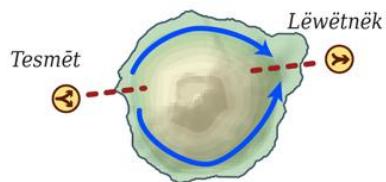
Map 5 - UP[wind] directions on **Motalava**



Map 6 - UP[wind] directions on **Gaua**



Map 7 - UP[wind] directions on **Vanua Lava**

Map 8 - UP[wind] directions on **Mota**Map 9 - UP[wind] directions on **Merelava**

In spite of their apparent simplicity, each of these maps constitutes a synthesis of a large number of data points which I accumulated over the years. The collection of this sort of navigational data often involved walking along the paths of these various islands, and asking my companions for the directions of different villages from various points in each island; I would also pay attention to the directionals people used as they mentioned other villages – as in ex.(4) above – or retold their own journeys walking on the island.

Crucially, in all islands of the Torres-Banks group I observed the same pattern. Along a large portion of the coast – about half of the shoreline, on each side of the island – the assignment of cardinal directions is constant: that is, if one goes ‘up’ from point A to point C, then any point B located between A and C will follow the same direction A→B→C. Thus on Gaua island (Map 6), if I left cape Vatlēs in the north and walked east along the coast, I would be going ‘up’; even after crossing the island’s northeast cape, I would still go ‘up’, during all my journey along the eastern shore. For long portions of the shoreline, this configuration results in an apparent system whereby one directional is constantly used when walking “clockwise” around the island (i.e. walking along the coast with the sea on one’s left-hand side), whereas the reverse directional would define a motion “counterclockwise” (i.e. walking with the sea on the right). For a casual observer, this configuration may be interpreted along the lines of “circular systems” which have indeed been reported for some islands of the Pacific, and which are based on a *clockwise–counterclockwise* contrast (see Lichtenberk 1983 for Manam, Dixon 1988 for Boumaa Fijian). Such systems exist, even though they are very rare (François 2004:15-16).

In fact, none of the spatial systems used in the Torres-Banks islands involves any circular coordinates properly speaking. While it is true that the directionals *up* or *down* will be used as an equivalent of “clockwise” for a large portion of the shoreline, there will always be a point (or rather, two points on each island) WHERE THE ORIENTATION IS REVERSED. For example, if I walk clockwise along the eastern coast of Gaua, I will be walking ‘up’ (i.e. ‘upwind’), until I reach a hamlet called Atkor. As I walk past that particular point and keep following the trail ahead of me, suddenly the direction I am following is not ‘up’ any more, it has become ‘down’ (see path on Figure 7). Indeed, villages on the south coast of the island point to Atkor as being located ‘up’ (Dorig *sag*); villages on the east coast also point to the same place using the same directional ‘up’. This means that Atkor is what I would call a FOCUS OF CARDINAL CONVERGENCE, i.e. a location on the island in which two instances of the same cardinal direction meet. Figure 7 illustrates the configuration characteristic of such a focus of convergence: two directionals referring to the same cardinal coordinate (here *up* for ‘southeast’), yet adapted so as to fit onto two different portions of the coast, converge at a particular point.

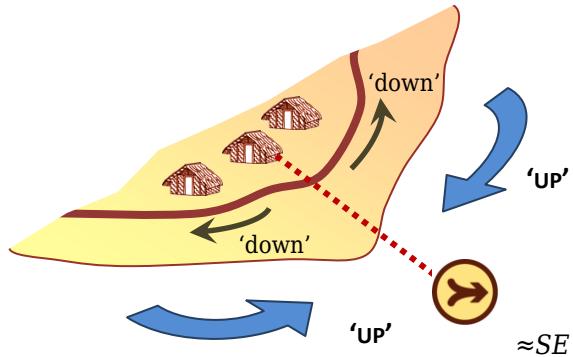


Figure 7 - A focus of cardinal convergence involving two UP vectors

Atkor, for example, is a place where two UP directions **converge**; by the same token, it is also a place where two DOWN directions **diverge**. Thus starting from the hamlet of Atkor, walking along the coast northwards (counterclockwise) is going *down* (Dorig *van sōw*), but walking southwards (clockwise) is also going *down*. At the other end of Gaua island, Cape Vatlēs offers the mirror configuration: it is the source point from which two cardinal UP vectors diverge - one running clockwise, the other counterclockwise; but also, logically, it is the target point where the two cardinal DOWN vectors converge.

The maps proposed above make use of two adhoc symbols:

- focus of convergence of two DOWN cardinal directions;  
= focus of **divergence** of two UP directions
- focus of **convergence** of two UP cardinal directions;  
= focus of divergence of two DOWN directions

For the sake of consistency, I follow everywhere the arbitrary convention of choosing the UP direction as the reference. This is why the first of these two icons represents *divergence*, and the second one represents *convergence*.

From the local perspective - e.g. that of an individual standing on the beach, facing the sea - these foci of cardinal convergence constitute paradoxical concepts, as they entail that two opposite directions along the coast will make use of the very same cardinal directional. This is exotic to outsiders, who would have a hard time imagining a place where 'south' would designate two exactly opposite directions (except perhaps on the North Pole). But it is also a paradox for the islanders themselves, who are sometimes confused, or amused, by these tipping points where both directions along the coast are 'up', instead of the normal situation in which one way is 'up' and the other way 'down'.

Paradoxical though these foci of convergence may be, their mechanism results logically from the combination of two elements: a single cardinal axis with a binary contrast UP-DOWN; and a principle specific to its use on land, whereby this axis must be redesigned so as to always run parallel to the shore. Because the cardinal axis is always adapted to *both* sides of the island, there are necessarily one source point and one target point for cardinal UP on land.

### 3.4.3 Plotting foci of cardinal convergence on maps

As Maps 2-7 showed, each island of northern Vanuatu, with no exception, is characterised by exactly two foci of cardinal convergence: one focus of convergence for UP vectors, and one focus for DOWN vectors (i.e. focus of divergence for UP vectors). The nature of these foci of

convergence can be quite diverse: some refer to capes and promontories (e.g. *Wotgrave*, *Ñerē Vot*, *Ñeyē Vēt*, *Ñus Ñereqō*...), others to hamlets or villages (e.g. *Liwotqeī*, *Tesmēt*, *Loli*, *Qēgmagde*, *Pāk*...).

The precise location of these tipping points is not entirely straightforward or derivable from a simple rule; in order to chart them, I had to perform careful checkings in each island separately, often in conjunction with a meticulous survey of the toponymy. On each of the maps above, the only placenames I indicate are the foci of cardinal convergence themselves; this indication should allow future researchers to retrieve them in the field and chart them with more geographical precision.

In spite of a certain amount of diversity visible from the maps, a general pattern emerges. On average, foci of **divergence** for UP vectors, symbolised by , are typically located on the NORTHWEST tip of their island; foci of **convergence**  are typically located on the SOUTHEAST tip. Of course this observation is expected, since I presented these directions ultimately as instances of an underlying NW-SE cardinal axis that has been mapped onto the shape of each island. However, the connection is more obvious in some islands than on others. The systems found on Hiw, Lo, Tegua, Mota or Gaua are all “well-behaved”, as it were, because the mean vector defined by the two foci of convergence (cf. dotted lines) aligns quite well with the NW-SE direction used on the navigational scale (cf. Figure 6 above).

Conversely, several islands have a global shape which lends itself less clearly to a NW-SE orientation; this results in distortions of the axis in ways which can be extreme, and occasionally render the interpretation less evident. For example, the first island I explored was Motalava (Map 5), a place where the straight line between the two foci happens to be oriented (S)W to (N)E. Knowing that the cardinal axis used across islands (cf. Figure 6 above) has quite a different orientation N(W)-S(E), it took a long investigation to understand that these two *up-down* axes, even though they run almost orthogonal to each other, were really two surface realisations of the same underlying axis (François 2003:426-434).<sup>13</sup> A similar difficulty would be represented by the use of *up* and *down* on the island of Toga – even though the same language (Lo-Toga) behaves much more classically on the neighbouring island of Lo.

All in all, the geographical location of foci of cardinal convergence follows a simple rule. On the one hand, the default location of the DOWN focus  and the UP focus  is, respectively, at the NORTHWEST and SOUTHEAST tips of an island. This is typically true of round islands (Mota, Gaua, etc.), whose circular shape favours the default case. On the other hand, if the geography of the island displays a prominent feature (cape, promontory) that is not exactly located on the NW-SE axis, then this feature tends to attract the focus of convergence, resulting in its apparent skewing with respect to compass terms. This is visible on Merelava, where the eastern cape of Lëwëtnök skews the axis towards ENE; on Ureparapara,

---

<sup>13</sup> Hyslop (2002:64) reports a similar mismatch for the island of Ambae further south (see Figure 6): “the absolute distinction made when describing motion within the island [is essentially SW vs NE]. Note that the distinction made for travel between islands is on a different axis, with islands to the south and east distinguished from those in the north and west.” While she describes this mismatch as “a curious variation in the division of absolute direction”, I believe it reflects a situation very similar to the case of Motalava, with ultimately the same underlying NW-SE cardinal orientation. This is confirmed by the existence – apparent from her map of Ambae (2002:62) – of a focus of cardinal UP convergence towards Lolosangga, on the middle of the eastern coast of the island.

where the southeastern cape of Nēyē Vēt has locally forced the cardinal axis almost into a NE direction; and so on. As a rule, though, the rotation of the axis never shifts more than 90° away from the underlying SW-NE cardinal axis (François 2003:431).

### **3.4.4 A note on standardised representations of geocentric systems**

A short note may be useful about my choice to represent all directional systems, in this article, using a standard representation such as in Figure 4 p.17. In order to facilitate comparison across systems, such figures follow the arbitrary convention of always representing NW on the left and SE on the right: that is, the figure will always represent an island seen from its western coast, regardless of the actual location of the speech community under discussion. For example, Figure 2 for Gaua languages gave a realistic account of the system used in the language of Lakon, because Lakon happens to be spoken on the western coast of Gaua: as Map 6 showed, this is an area where ‘up’ is used for “counterclockwise” (i.e. walking along the coast with the sea on one’s right handside).

As for the language of Nume, which is primarily spoken on the northeast coast of the island (Map 1 p.7), a realistic representation of the way in which it is used in its native area would normally require reversing Figure 2, with SE (‘up’) pointing left and NW (‘down’) pointing right. However, this increase in realistic accuracy may result confusing for the comparison of similarities and differences across languages. I prefer to adopt everywhere a standardised view of a fictional island seen from its western coast, for all languages – even when their community is actually located on an eastern coast. One reason for considering this decision legitimate is the fact that language speakers are geographically mobile, and regularly have to adapt their own space system to other environments, even outside their home village (cf. François 2003:428). In this sense, a representation such as Figure 2 portrays accurately any language, not just those that are typically spoken on a western coast.

## **3.5 Summary**

Section 3 was concerned with the way in which the 17 Torres-Banks languages encode absolute directions on the navigational scale, i.e. for space reference involving longer distances. In Table 6 p.16, this corresponds to rows #8–9 and #12–13.

In spite of apparent or superficial differences, the general observation was one of a profound homogeneity – as far as this scale is concerned – among all languages of the area. Everywhere, languages make use of a single cardinal axis oriented NW-SE for their navigation across islands. Everywhere, this axis is lexified using the vertical directionals *up-down* – either synchronically, or historically. This *up-down* cardinal axis is evidently inherited from the earlier Oceanic system, and ultimately finds its source in the seafaring traditions of Austronesian navigators, for whom southeast trade winds provided the basis for a contrast *up(wind)* vs *down(wind)*. Finally, we saw that the same languages also use this cardinal axis on land, at least for long distances, e.g. when referring to other villages on the same island. The mapping of the cardinal axis onto the shoreline results in the paradox of *foci of cardinal convergence* – always two per island – in which identical directionals meet. The geography of particular islands occasionally impose on the axis some amount of skewing ( $\pm 90^\circ$ ) in compass terms, yet it can be shown that the underlying orientation of the axis remains everywhere NW-SE.

As we shall see in the next sections, some languages reserve the NW-SE cardinal axis to longer distances, while others even employ it for shorter distances. This belongs to the domain of the *local scale*, where more cross-linguistic diversity can be found.

## 4 Two canonical systems and their variations

The remainder of this survey will concentrate on the geocentric directionals as they are used on land, and on the local scale. This is the domain where the nine systems differ most, which is why each system will be presented separately. In doing so, I shall follow a form of logical order, going roughly from the simplest to the most complex languages of the area.

The present Section 4 will start with two simple systems: Gaua and Mwotlap. The one used in the four indigenous languages of Gaua (already illustrated for Dorig in §1.3) is the most conservative. While it can be seen as equally simple in its make-up, Mwotlap reflects a number of innovations that I will unravel. I will then propose (§4.4) that Gaua and Mwotlap can be taken as two “canons” with respect to which the more intricate space systems of northern Vanuatu can be understood. Among these, Section 5 will examine Volow and Vanua Lava, two systems which are non-canonical mostly on their coastal axis. Section 6 will describe Löyöp, Mota and Mwerlap, three systems whose peculiarities lie mostly on the land-sea axis. Finally, Section 7 will describe the more complex system shared by Lehali and Lo-Toga, before Section 8 attempts to unravel the quirkiest of all languages in the region: Hiw.

### 4.1 The local scale: presentation

The local scale refers to spatial reference within one's sight and action range – typically for distances shorter than one or two hundred meters. In all Oceanic languages, this local scale involves two orthogonal axes: on the one hand, a *land-sea* axis that runs from the centre (or central mountain range) of the island down towards the sea, and often further towards the deep ocean; on the other hand, a secondary axis running parallel to the shore, which I will call the ‘coastal axis’.

Northern Vanuatu languages are no exception to this Oceanic prototype. The principal axis used on land defines everywhere a contrast ‘inland’ vs ‘seawards’, and runs from the island's watershed all the way to the shore. In all languages, the directionals used on dry land – e.g. in the village – can also function at sea: for someone on a canoe, paddling ‘inland’ involves coming closer to land, whereas paddling ‘seawards’ means going further away towards the ocean. This pattern is visible from Table 6 p.16: everywhere, the directionals for ‘landwards’ and ‘oceanwards’ at sea (rows #4 and 17) are identical with those contrasting ‘inland’ and ‘seawards’ on land (rows #5 and 16).

Languages do differ, though, in the way they lexify this land-sea axis: some use a contrast *up-down*, others a pair of directionals *in-out*. They also differ in how they express the axis that runs orthogonally to that land-sea axis, i.e. parallel to the shore: some systems here use the cardinal axis we saw earlier (§3.4), others don't. Finally, some languages have developed different subsystems depending on distance contrasts, or on location (in the village vs in the bush). I will now go through each of the nine systems one after the other, in order of increasing complexity.

## 4.2 A conservative system: Gaua languages

The simplest system of absolute reference found in the Torres-Banks group is no doubt the one used on the island of Gaua. As Table 6 showed, exactly the same structural organisation is found in the five indigenous<sup>14</sup> languages of this island: Nume, Dorig, Koro, Olrat, Lakon. This system is represented on Figure 8 – repeated from Figure 2 above.

### 4.2.1 Two distinct up-down axes

An analysis of Figure 8 shows that Gaua languages use vertical directionals *up-down* for geocentric reference, with two very different meanings. First, these define the cardinal axis used on the NAVIGATIONAL scale (§3.4), whether across islands (cf. rows #8+13 in Table 6) or across distant villages spoken on the same island (rows #9+12); as we know, this *up-down* pair originates historically in a contrast *upwind* vs *downwind*. But the same languages also use the vertical directionals on the LOCAL scale to encode the land-sea axis, whether in the bush (rows #6+15) or on the coastal area where most villages are (rows #5+16); as mentioned above, the contrast *up-down* extends even at sea along the same axis, to contrast ‘towards the island’ with ‘towards the open sea’ (rows #4+17). The motivation for the use of vertical directionals along that land-sea axis obviously originates in the literal use of ‘up’ and ‘down’ in the hilly areas of the island: *up-down* here can be glossed *uphill* vs *downhill*. The vertical contrast has simply been extended to the flatter areas of the island, and even to the horizontal plane of the lagoon, as these directions came in continuity with the more salient slopes from the island’s central mountains.

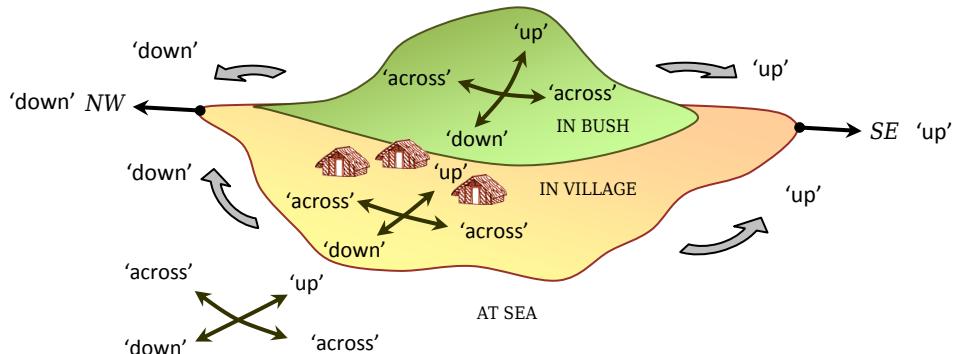


Figure 8 - The system of geocentric directionals in Gaua languages

The following examples, taken from various Gaua languages (Lakon, Nume, Dorig), illustrate the use of *up* and *down* for the land-sea axis:

- (LKN.24) We gēē tē van *ajew hag* le vanō.  
and 3pl SEQ go upwards up LOC village  
(They landed on the shore) and began to walk *up* towards the village. [LKN.Qat:060]
- (NUM.25) Bas nen, ni tov rev *tēqēl* wak *ror* le won.  
end TOP 3sg SEQ pull downwards canoe down LOC sand  
‘Then he dragged the canoe *down* to the beach.’ [NUM.d07:13]

<sup>14</sup> The community of about 400 Mwerlap speakers who have settled on Gaua use a different system, adapted from the one used on their home island: see §6.3.

- (DRG.26) Nēk so sō *swēl* lala mlē **ror** le lam ni.  
 2sg POT<sub>1</sub> paddle downwards POT<sub>2</sub> also down LOC ocean INSTR  
 'You can even paddle it further (*down*) towards the ocean.'

[DRG.d07:14]

In the language Lakon, the directionals *hag* vs *hōw* have a free variant, respectively *roka* 'up' and *rōkōw* 'down':<sup>15</sup>

- (LKN.27) Nē 'n gih nē tē ol mē lē umā **rōkōw** rek.  
 3sg PFT seize 3sg SEQ enter with.it LOC house down DIST  
 'He grabbed (the knife) and brought it in that house *down* there.' [LKN.d05:20]
- (LKN.28) Magte neñ ēn hag suu *hōw* **roka** neñ a Liwsal.  
 old.woman MED PFT sit downwards down up MED LOC Liwsal  
 'The woman sat down in the river, *up* there in Liwsal.' [LKN.Origin.Lake.16]

The two *up-down* axes run orthogonal to each other, and are obviously distinct. Occasionally, the homophony may trigger some confusion: for example, *sō ror* 'paddle down' can mean 'go out towards the ocean' as in (26), but it can also mean 'travel on a canoe towards northwest' (cf. Figure 6 p.20). Systems involving two distinct *up-down* axes have been reported for other Oceanic languages (Hyslop 2002; Palmer 2002:128); and this is in fact the system that has been reconstructed for Proto Oceanic (François 2004:21). In all systems showing that configuration – including those found on Gaua – the risk of confusion between scales is somewhat limited by the fact that, in principle, the two *up-down* axes never really cross: the wind-based contrast belongs to the NAVIGATIONAL scale, whereas the axis based on the declivity of the ground pertains to the LOCAL scale. For directions parallel to the shore involving short distances, Gaua languages never use their cardinal coordinates. Instead, they resort to an undifferentiated traverse axis—the topic of next section.

#### 4.2.2 The undifferentiated traverse axis

When speakers of Gaua languages, on the local scale, refer to a location that is neither 'inland' nor 'seawards', they use a directional that belongs to an axis orthogonal to the main land-sea axis ("traverse"), and therefore parallel to the shoreline.

Crucially, this coastal axis uses the same directional on both sides: this is visible from Table 6, which shows that Gaua languages use the same form for #10 '*parallel to the shore towards SE (close)*' and #11 '*parallel to the shore towards NW (close)*'. I thus describe the axis as "undifferentiated", using the term proposed by Palmer (2002:127): "a derived axis for which a language does not lexically distinguish the opposing directions". In a way, this configuration whereby two opposite directions along the coast use the same directional is reminiscent of the cases of *cardinal convergence* we saw in §3.4.2. The difference is that foci of convergence constituted isolated exceptions within a more general rule of differentiation *up-down* on the navigational scale; whereas it is an inherent property of the traverse axis to be undifferentiated, wherever it is used on the island.

A convenient gloss for this directional is 'across', as it crosses the primary axis. Etymologically, some of these forms (Nume *van*, but also Mota *vano* or Mwerlap *van*) reflect a

<sup>15</sup> Notice, in (28), the rare combination of two opposite directionals in the same clause: *hōw* referring to the motion of the event ('sit down'), and *roka* locating that event in the geography of the island ('up there inland').

POc verb *\*pano* ‘move in transverse direction’ (cf. Figure 5 p.19); other forms (*vak*, *pāh*) reflect local innovations (§11.3.3).

- (DRG.29) Dār s-van *barbar* ***vak*** seg dār s-ñor ***vak*** sa?  
 1INC:du IRR-go crosswise across here 1INC:du IRR-sleep across there  
 ‘Why don't we walk (across) over there and have a nap?’ [DRG.Heron.17]
- (LKN.30) Miini neñ, tē van *päätäg* gēn ***päh***.  
 child that PRSTV go crosswise FOC:there across  
 ‘That child, there he is, heading over there (*along the beach*).’ [LKN.d05:34]

We saw earlier (§2.4.2) that directionals are sometimes reinforced by a postverbal POST-VERB with a similar meaning. Interestingly, the traverse directional glossed ‘across’ is often associated with an postverb that translates as ‘crosswise, orthogonally’ (as in ‘lie across the bed’); this confirms the choice of glossing the directional ‘across’. Table 7 provides the forms of the postverbs with directional meanings (as distinct from *directionals* strictly speaking) for the five languages of Gaua. The reader will recognise some postverbs found in (24)–(26) above.

Table 7 — Directional postverbs (distinct from directionals) in Gaua languages

	Nume	Dorig	Koro	Olrat	Lakon
‘upwards’	kal	kal	kal	kal	ajew
‘downwards’	tēqēl	swēl	sēwēl	sēw	sēwēl
‘crosswise’	barabar	barbar	bertäg	paytaa	paatäg

#### 4.2.3 A conservative system

The undifferentiated traverse axis is only used on the local scale, for distances shorter than about 100 or 200 m: this corresponds more or less to the array of what is in sight of the speech act participants, and typically matches the size of the village. For longer distances on land, directions parallel to the shore make use of the cardinal axis, as we saw in §3.4. At this scale, the coordinates on each side become differentiated again, in the form of another contrast *up-down*. This cardinal axis is typically used for distances across villages, out of sight, along the coast.

Figure 8 above represented the various vectors used in the languages of Gaua. The discussion above has established that some of these vectors follow a single logic: for example, the use of DOWN at sea (‘towards deep sea’) is merely an extension, on the horizontal plane, of the underlying meaning *down(hill)*. In other words, the vectors #15-16-17 of Figure 4, which are lexified separately in some other languages (§6), are colexified in the languages of Gaua, and treated as three instances of a single emic category. Put together, the various emic categories of Gaua directionals make up the system represented on Figure 9.

The system used in Gaua involves only three axes: one cardinal axis *up(wind)-down(wind)*; one topographical axis *up(hill)-down(hill)*; and one undifferentiated traverse *across*. Crucially, this is exactly identical to the system that has been reconstructed for Proto Oceanic (François 2004:16-18). In other words, the geocentric system used today on Gaua is highly conservative

of its ancestor.<sup>16</sup> This is one example, among many others, where languages can prove conservative in their structures, even though the lexical material used to lexify them may itself have gone through local innovations (François 2010, 2011:226).

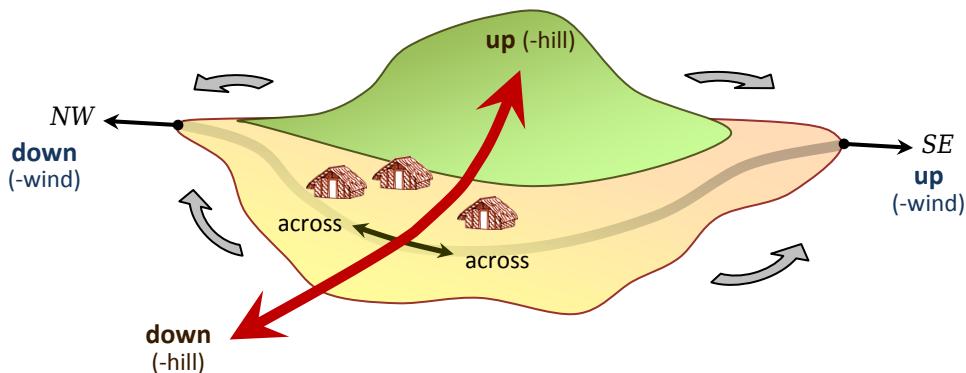


Figure 9 – The system of geocentric directionals in Gaua languages (emic representation)

The languages of Gaua provide a good point of departure for the next sections, which will examine successively the other directional systems found in the region. By contrast with the simple and conservative system of Gaua, all these systems can be characterised by a number of local innovations, each of which has tweaked the original configuration of the ancestor. The accumulation of these innovations has resulted in the wide diversity we can observe today.

#### 4.3 A complete redesign: Mwotlap

Compared to Gaua, a language like Mwotlap, spoken on Motalava island, shows quite a different system. The only agreement between Mwotlap and Gaua – as well as all other languages of the region, for that matter – is the use of an *up-down* cardinal axis coding for a NW-SE contrast on the navigational scale (§3). But all other features of the system are different.

First, Mwotlap has lost the undifferentiated transverse axis ('across') that is found on Gaua, and reconstructed for Proto Oceanic. Directions parallel to the shoreline are encoded using the cardinal axis *up-down* everywhere on land, not only for long distances across villages (§3.4), but even for short distances, on the local scale. Sentences like (1)-(2), repeated here, would be impossible in Gaua languages, yet are perfectly common in Mwotlap:

- (MTP.1) na-bankēn mey hag tō lok **hag**  
ART-mug REL sit PRSTV side up  
'the mug on the [south]east side (of the table)'

- (MTP.2) Kē mi-tig lō-tōti beg, ba lok **hōw**.  
3sg PRF-stand LOC-trunk breadfruit but side down  
'She's standing at the breadfruit tree, on the [north]western side.'

<sup>16</sup> We already saw in §2.4.2 that the same Gaua languages are also conservative of the Proto Oceanic subsystem of topological directionals.

The second major difference between Mwotlap and Gaua is that the land-sea axis is never encoded by the vertical directionals *up-down*, but by a contrast between ‘in’ and ‘out’ (cf. Table 5 in §2.4.2). Sentence (31), taken from a traditional story, takes place as the main character Venventey, living on a coastal village, comes down to the beach to welcome his brother who’s arriving on a canoe. Out of the six directionals used here, three refer to the land-sea axis: first as Venventey walks down to the beach, second as they both carry the canoe to the beach, before finally walking up to the village:

(MTP.31)	Kē	ni-van	<b>yow</b>	tō	ni-tēy	<b>van</b>	ni-siok	nonon	tō,
	3sg	AO-go	<b>out</b>		then	AO-hold	thither	ART-canoe	his
	kōyō	hah	kal	<b>hay</b>	tō,	leveteg	<b>van</b>	lē-vēthiyle.	
	3du	lift	upwards	<b>in</b>	then	put.down	thither	LOC-sand	
	Kōyō	hatig	<b>hag</b>	tō,	van	<b>hay</b>	l-ēm	ēgēn.	
	3du	rise	up	then	go	<b>in</b>	LOC-house	now	
	‘So he walked down to the shore [lit. went <b>out</b> ], took hold of his canoe; they both carried it up <b>in</b> (land), and put it down on the sand. Finally they left the place and walked <b>in</b> (land) towards their house.’								

[MTP.Venventey.WS.072]

Mwotlap cannot use ‘up’ and ‘down’ here, but resorts to **hay** ‘in’ and **yow** ‘out’. Whereas Gaua languages, on the local scale, use ‘up’ and ‘down’ for an axis orthogonal to the shoreline (*uphill* vs *downhill*), Mwotlap uses them exclusively for those directions which are parallel to the shore: these correspond to the cardinal axis (originally *upwind* vs *downwind*) once it has been mapped onto the land (§3.4) and generalised to all scales. The most likely scenario is that the extension of the cardinal axis to the local scale would have resulted in a conflict between two orthogonal *up-down* axes; the potential misunderstandings that would have necessarily resulted from this innovation was evidently solved by a second innovation, namely the relexification of the land-sea contrast using the topological directionals *in-out* (François 2004:22-26).

This lexical choice represents the whole island as if it were a container: walking away from the sea into the more bushy areas of the island is going ‘in’ (cf. Eng. *inland*) whereas walking away from the bush and towards the sea is equivalent to going ‘out’ (cf. Eng. *out to sea*). Mwotlap keeps the same contrast at sea. As long as a landmass is salient to the observer, the land-sea contrast will be encoded as *in-out* – even when referring to a shoal of fish in the water:

(MTP.32)	No	m-et	nō-mōmō	ni-sey	<b>hay</b>	ni-sey	<b>yow</b> .
	1sg	PFT-see	ART-fish	3sg-move.in.shoal	<b>in</b>	3sg-move.in.shoal	<b>out</b>
'I saw a shoal of fish moving <i>in</i> (landwards), and suddenly moving <i>out</i> (oceanwards).'							

[BP5-34a]

Overall, these two innovations (generalisation of the cardinal axis to the local scale; relexification of the land-sea axis as a consequence) have resulted in a newly designed system. By comparison with the two *up-down* axes found in conservative languages, and the third transverse axis, a system like Mwotlap only involves two axes: one cardinal axis lexified *up-down*, employed everywhere on land for vectors parallel to the shoreline (see Map 5 p.23); one land-sea axis running orthogonal to it, lexified *in-out*. The geocentric system of Mwotlap – described in more detail in François (2003) – is represented in Figure 10. It is, arguably, the simplest system among all northern Vanuatu languages.

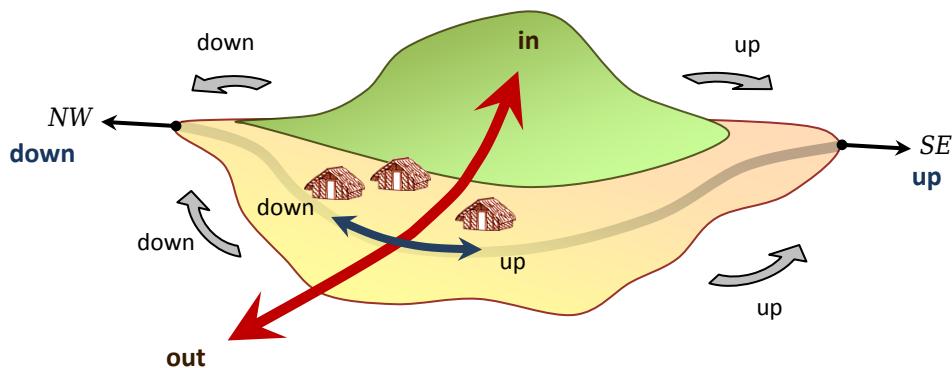


Figure 10 - The system of geocentric directionals in Mwotlap

#### 4.4 Two canonical systems and a number of hybrids

The Mwotlap system of orientation, with its two orthogonal axes (Figure 10), represents a completely new design compared with the conservative system found on Gaua (Figure 9). In the remainder of this paper, I will propose that the observation of space systems of northern Vanuatu can be accounted for by positing Gaua and Mwotlap as two opposite “canons”—to use a terminology inspired from Corbett’s (2007) *canonical typology*—each of which displays a coherent set of space-related properties.

By comparison with these two canons, the other languages of the region present hybrid systems, i.e. systems which are closer to one of the two canons, yet deviates from it in ways that make it resemble the other canon. For example, we will see that Mota has almost exactly the same system as Gaua, except that it uses a directional ‘out’ when pointing seawards, in a way similar to Mwotlap. Symmetrically, Löyöp is almost like Mwotlap, except that it still uses the *up-down* contrast (like Gaua) in the steeper areas of the island. Because the comparison between systems involves several parameters, it is not possible to rank them using a unidimensional scale, whereby languages would simply placed in a linear order between the two poles Gaua and Mwotlap. Instead, we are dealing here with two canons, from which each language deviates following a number of dimensions.

Derived from the data in Table 6 p.16, Table 8 lays out the main relevant parameters whereby languages differ in their directional systems.

Table 8 – The systems of Gaua and Mwotlap constitute two canons; all other northern Vanuatu systems can be analysed as hybrid or intermediate between these two.

System	PARALLEL TO SHORELINE					ORTHOGONAL TO SHORELINE			
	navigational		local up-down		traverse	up-down		in-out	
	‘up’	‘down’	‘up’	‘down’	‘across’	‘up’	‘down’	‘in’	‘out’
<b>Gaua lgs</b>	+	+	—	—	+	+	+	—	—
Mota	+	+	—	—	+	+	(+)	—	(+)
Mwerlap	+	+	—	—	+	(+)	(+)	(+)	(+)
Vanua Lava lgs	+	+	—	+	(+)	—	—	+	+
Volow	+	+	+	+	—	(+)	—	(+)	+
Löyöp	+	+	+	+	—	(+)	(+)	(+)	(+)
Lehali, Lo-Toga	+	+	+	+	—	(+)	(+)	(+)	(+)
Hiw	—	+	—	+	—	+	(+)	—	(+)
<b>Mwotlap</b>	+	+	+	+	—	—	—	+	+

From the first to the last column, the relevant parameters can be defined as follows:

1. for directions PARALLEL TO THE SHORELINE:
  - whether the navigational subsystem used for long distances on land (§3.4) employs the cardinal directionals ‘up’ [#9] and ‘down’ [#12];
  - whether the local subsystem used for short distances [#10, #11] employs those same cardinal directionals ‘up’–‘down’, or an undifferentiated traverse;
2. for directions ORTHOGONAL TO THE SHORELINE:
  - whether the land-sea axis [#4–5–6] employs the vertical directionals *up-down* or *in-out*.

Table 8 uses brackets “(+)” to indicate those cases where the answer to these questions is not straightforward, or depends on certain conditions. For example, we’ll see that Mota uses sometimes *down* when pointing to the sea, and sometimes *out*, depending on how steep the slope is; likewise the languages of Ureparapara and the Torres Is encode the land-sea axis as *up-down* in the bush, but as *in-out* in the lower parts of the island.

To take another example, we’ll see that the languages of Vanua Lava employ a directional that can be glossed ‘across’ for directions parallel to the shoreline, yet instead of having it for both directions as in Gaua, they only use it for *one* direction, towards SE. This, again, is an example of a transitional configuration between the conservative languages of the southern Banks Islands (including Gaua) – where the transversal directional ‘across’ is used in both directions – and the innovative languages of the northern Banks and Torres (including Mwotlap) – where no such directional exists at all.

Other examples of hybrid configurations will be detailed in the sections below. In almost all cases – except for some peculiarities of Hiw – the languages of northern Vanuatu can be shown to pattern partly like the conservative system of Gaua, and partly like the innovative system of Mwotlap.

## 5 Variations on the coastal axis

### 5.1 Accidental homophony: Volow

I will briefly examine the special case of Volow, a communalect spoken on the eastern side of Motalava island, where Mwotlap is also spoken. Even though Volow is now quasi extinct, some valuable narratives were recorded in 1969 by the anthropologist Bernard Vienne with Wanhan, the last fluent speaker of the language; I transcribed them in 2003 with the help of Wanhan’s sons.

The data from Volow in Table 6 p.16 show a threefold organisation of geocentric directionals: the two forms **hō** ‘down; northwest along the coast’ and **yo** ‘out; seawards’ stand in contrast to a single polysemous directional **ha** ‘up, southeast; in, inland’. The ambiguity of the latter form can be represented with a gloss ‘up/in’. This pattern of colexification ‘up/in’ is also found in the Volow system of topological directionals (§2.4.2), and is therefore imported into the geocentric system. Figure 11 lays out the system of Volow.

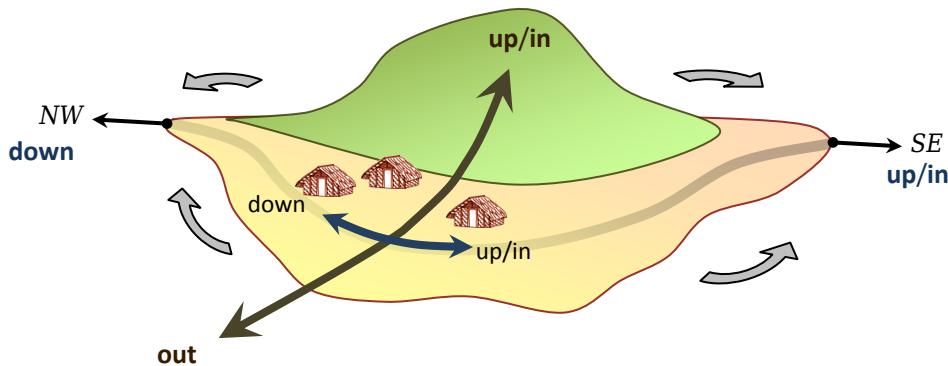


Figure 11 - The system of geocentric directionals in Volow

Those who can understand Volow recordings today point out systematic correspondences between their ancestors' space system and the one of the dominant language Mwotlap, which they have now shifted to, and which I used as my contact language. Thus, in spite of the apparent ambiguity of the two *ha* directionals in (33), these speakers can associate each token with the corresponding directional of Mwotlap. The first *ha* is here glossed 'in, inland' (MTP *hay*), and the second one is 'up, southeast' (MTP *hag*):

- (VLW.33) N-tēqē mine yiwas *ha* qe gēs, taval tō, teyeg *ha*.  
 ART-garden my close *up/in* FOC here beyond hill side *up/in*  
 'My garden is close this way (*inland*), it's on the other side of the hill,  
 towards *southeast*.'

[VLW.d01:16]

The colexification of 'up' and 'in' is not unique to Volow, and we saw it also in Gaua languages. Synchronously, it results in the same sort of ambiguity, whereby the same form is used on two different axes. These ambiguities, however, probably have different explanations. In the case of Gaua, the pattern was evidently conservative of Proto Oceanic \**sake*, which was already ambiguous between a vertical use 'up' and a topological use 'in' (see §2.4.2). In the case of Volow, by contrast, it is likely that the colexification results historically from the more recent merger of two erstwhile distinct directionals. Considering how close Volow is from Mwotlap in all other respects (Kalyan & François f/c), it is likely that Volow once had the same four-member system as Mwotlap – see Figure 10 p.34. At some later point, all directionals in Volow underwent the deletion of their last consonant (\**hōw* → *hō*; \**yōw* → *yo*; \**vān* → *va*); this resulted in the accidental homophony of two directionals *ha*, one meaning 'up, southeast' (< \**hag*), the other meaning 'in, inland' (< \**hay*).

## 5.2 An incomplete redesign: Vanua Lava languages

Vanua Lava is the largest and the tallest of the Torres and Banks islands (§2.1). If one sets apart the relatively recent colonisation by Mwotlap speakers on its northeastern coast, Vanua Lava is home to four distinct languages: Vurēs, Vera'a, Mwesen and Lemerig. Despite their numerous differences, these four languages share an identical system of space reference. Just like Gaua languages, those of Vanua Lava can therefore be described as a single geocentric system since they employ the same emic categories – even though the precise phonological forms assigned to each category differs across languages.

### 5.2.1 An asymmetry on the transversal axis

In a nutshell, the systems of Vanua Lava are identical to the one we saw for Mwotlap (Figure 10 p.34), except for a single vector: the one that points southeast for directions parallel to the shore, on the local scale. On the one hand, long distances along that vector are encoded with the vertical directional ‘up’ (*siag* in Vurës, *sag* in other languages): this simply reflects the mapping of the cardinal axis onto land, as illustrated by Map 7 p.23. However, unlike the canonical Mwotlap system which generalises this use of ‘up[wind]’ to all distances on land, Vanua Lava languages reserve it for long distances, and make use of a distinct directional for distances shorter than about 200 meters – see the forms in Table 6 p.16 (row #10).

The two following sentences, taken from narratives in Lemeric, illustrate the contrast between the two directionals pointing southeast along the coast: *sag* for long distances, *wël* for nearby locations.

(LMG.34)	É	Qet	tär	e	'og'og	<b>sag</b>	sā	Lēseper	ow.
	PERS	(hero)	3pl	DEF	IPFV~stay	(SE:far)	FOC <sub>1</sub>	L.	FOC <sub>2</sub>
'Kpwet and his [brothers] were living over there (southeast) in Leseper.'									

(LMG.35)	Ti	m-'är	pa'	<b>wël</b>	kē	ge	mälägläg.	
	3sg	PFT-stand	hidden	(SE:near)	place	STAT	dark	
'He stood hiding over there (southeast) in the dark.'								[LMG.Rock.048]

Oddly enough, this distance-based contrast only exists in one direction. As for the opposite direction, it employs the directional ‘down’ whatever the distance – just like Mwotlap. As an illustration, the following excerpt mentions the four directionals that constitute the local-scale subsystem of Lemeric: *wël* ‘parallel to the shore, SE side’ – *sōw* ‘down; parallel to the shore, NW side’ – *sar* ‘in; inland’; *row* ‘out; seawards’. Notice here the absence of *sag* ‘up’.

(LMG.36)	Ti	m-sék.	Sék lu	<b>wël</b>	nē,	sék lu	<b>sōw</b>	nē,
	3sg	PFT-seek	seek	around	(SE:near)	there	seek	around (down/NW) there
	sék lu	<b>row</b>	nē,	sék lu	<b>sar</b>	nē	—	ti   'esgö'   qäl'ä.
	seek	around	out	there	seek	around	in	there — 3sg find   NEG
'So he began to search. He searched southeast, he searched northwest, he searched seawards, he searched inland — but he couldn't find it.'								

[LMG.Qet.072]

The system of Lemeric is thus asymmetrical, in the sense that a pair of distinct terms used in one direction (*wël* vs *sag*) contrasts with a single term (*sōw*) on the opposite direction.

The three other languages of the island, namely Mwesen, Vurës and Vera'a, share the same properties as Lemeric.<sup>17</sup> The system of Vanua Lava directionals is shown on Figure 12. The gloss ‘across’ given for the short-distance vector pointing southeast will be explained in §5.2.2.

---

<sup>17</sup> One peculiarity of Lemeric is that the directional *wël*, which here encodes a geocentric direction pointing southeast, has also grammaticalised as a participant-based allocentric directional ‘thither, towards non-speaker’ (§2.4.1): the latter contrasts with *me* ‘hither’, and can point to any direction. Conversely, the cognate forms in the three other languages of Vanua Lava (*wōl*, *wol*) are restricted to their geocentric use.

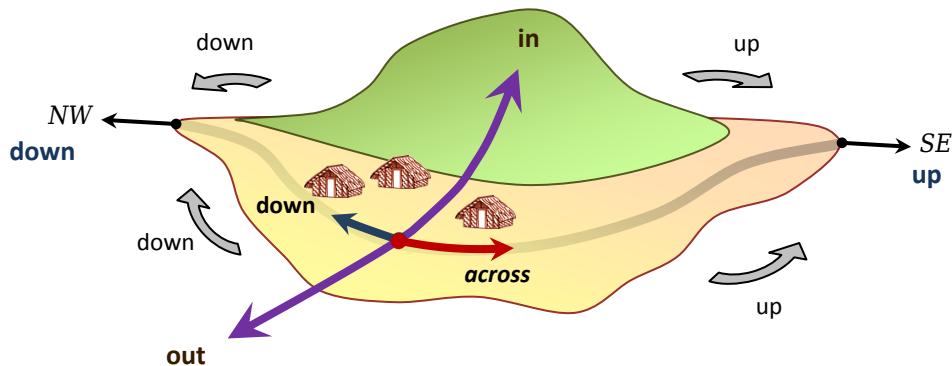


Figure 12 – The system of geocentric directionals in Vanua Lava languages

### 5.2.2 The ghostly memory of the lost traverse

The oddity of the Vanua Lava system can be interpreted by looking at the origin of the short-distance directional that contrasts with 'down'. In the language Vurës, the form *wōl* [wɔl] is homophonous with the noun for 'moon' *wōl* [wɔl], which might have suggested an etymological connection; however, examination of regular vowel correspondences reveals that the two words cannot be cognate. The modern forms – Vurës and Vera'a *wōl*, Mwesen *wol*, Lemerig *wēl* – demonstrably reflect a protoform \**volo*, whose meaning can be reconstructed as 'crosswise, across' (§11.3.3); cognate with it are such words as Mota *wolowolo* '[ADV] crosswise; [N] a cross' and Vurës *wōlōwōl* '[N] a cross; a crossbeam' (François f/c:195).

Interestingly, one finds here exactly the same meaning as the one associated with other, non-cognate spatial terms which are found in the languages of Gaua (§4.2.2). The main difference is that the latter languages use their 'crosswise' term – e.g. Dorig *barbar* – for both directions parallel to the shoreline, whether NW or SE; whereas Vanua Lava languages use it exclusively for one direction.

My hypothesis is that Vanua Lava languages once had an undifferentiated traverse axis similar to the one found in Southern Banks languages. While their ancestor must have lexified that axis using the POC directional \**pano* 'move in a transverse direction' (cf. Figure 5 p.19), events of lexical replacement resulted in the substitution of \**pano* with \**volo* 'crosswise'. This relexification was followed by a more important change, namely the generalisation of the cardinal axis 'up-down' on land, following the scenario outlined in §3.4 and 4.3. For languages like Mwotlap, the redesign was complete: the two 'across' vectors of the original traverse axis disappeared, and were replaced respectively by the 'up' and 'down' of the cardinal axis. However, for the four languages of Vanua Lava, the redesign was only partial: while their transverse axis 'across' was replaced by the cardinal 'down' on the northwest side, for some unknown reason it resisted attrition in the other direction. As a result, while Vanua Lava languages followed exactly the same innovations as Mwotlap, they have preserved the original distance-based contrast for directions pointing southeast.

If we compare Figure 12 with Figures 10 and 14 above, it becomes clear that Vanua Lava can be described as a hybrid between the canonical systems of Gaua and Mwotlap.

## 6 Variations on the land-sea axis

### 6.1 Depending on the slope: Löyöp

Löyöp, the language spoken on the eastern side of Ureparapara, shares with its neighbour Mwotlap the generalisation of the cardinal axis *up-down* on the local scale, for directions parallel to the shoreline. Löyöp also aligns with Mwotlap – or with Vanua Lava languages for that matter – in using the two directionals *in-out* to encode the land-sea axis:

- (LYP.37) Yege oñ e m-van me, m-van me, m-van me; m-ol kal  
 PL ship DEF PFT-go hither PFT-go hither PFT-go hither PFT-land upwards  
**say** me, lilwon, m-qēt me. Kyeyjöl m-van **yow.**  
**in** hither on.beach PFT-complete hither 3trial PFT-go **out**  
 'The ships kept coming closer, closer, closer, until they landed [*in*] on the beach, one after the other. The three boys walked down [*out*] towards them.' [LYP.Pig.117]

However, contrary to its neighbours, Löyöp reserves these directionals *in-out* to the flatter parts of its island, namely the coastal villages and the sea. By contrast, it employs the vertical directionals *up-down* in the forest and steeper parts of the island, where the declivity of the ground is more salient. The following sentence, taken from a traditional narrative, shows how Löyöp can use its vertical directionals to lexify the *land-sea* axis. The story mentions a hamlet located in the mountain:

- (LYP.38) Kyeyō m-yēm kal n-wutwut, van van van en:  
 3du PFT-climb upward ART-hill go go go TOP  
 kalō **sa** lipnō yo-yō.  
 arrive up in.village POSS-3du

'They climbed the hill all the way up, until they reached [*up*] their hamlet.' [LYP.Ogres.18]

The use of 'up' in such a context seems logical, and indeed would be the expected strategy in most Oceanic languages. However, the neighbouring languages Mwotlap or Lemerig would have to use 'in' in such a situation: they can only use their 'up' directional with its navigational use, on the horizontal plane. Löyöp, by contrast, has preserved the possibility to use its vertical directionals on the land-sea axis, as soon as the declivity of the slope warrants it.

The system of Löyöp is represented on Figure 13.

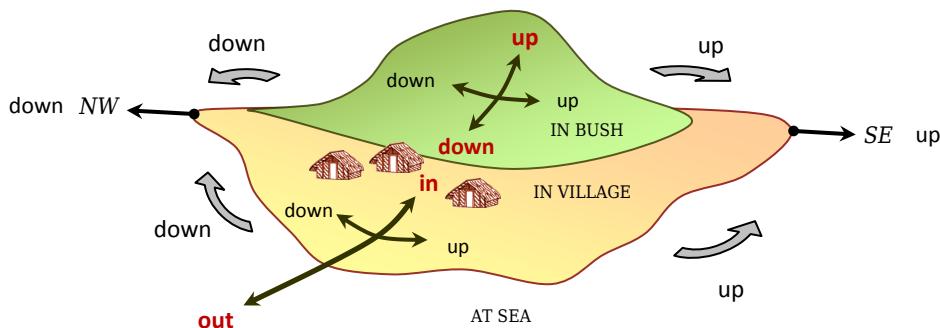


Figure 13 – The system of geocentric directionals in Löyöp

In sum, Löyöp is hybrid between the two canonical systems of Mwotlap and Gaua (§4.4). It resembles Mwotlap in the lower areas of the island, yet preserves some features of the ancient system in the heights, where two distinct *up-down* axes intersect. While Vanua Lava

had preserved an element of the ancient system on its transverse axis, Löyöp shows a similar form of conservatism, but this time on the land-sea axis.

The other systems that remain to be presented in this section all form, in turn, variations of the Löyöp case. Whether it is Mota, Mwerlap, Lehali, Lo-Toga or Hiw, all the languages yet to be examined show a similar lexical split within the land-sea axis, where they employ sometimes the vertical directionals *up-down*, and sometimes the topological *in-out*. They are the object of the next subsections.

## 6.2 An asymmetrical system: Mota

The small island of Mota, located east of Vanua Lava (see Map 1), uses a directional system that is essentially similar to the Gaua languages – see Figure 9 p.32. Mota uses two UP-DOWN axes: one corresponding to the land-sea axis (but see below), and one for cardinal directions mapped onto the shoreline (§3.4). Sentence (39) illustrates a dialogue that would take place in a western village of the island, such as Veverao, pointing towards the southeastern hamlet Liwotqei (see Map 8 in §3.4.2):

- (MTA.39)    Ka    vava    i    vea?    — Na    vava    iake    **sage**    a    Liwotqei.  
               2s:AO    IPFV~go    ALL    where                    1s:AO    go~IPFT    here    up    LOC (village)  
               'Where are you going? — I'm going *up* (southeast) to Liwotqei.'

[FP1-40b]

On the local scale, the land-sea axis is crosscut by an undifferentiated transverse axis lexified *vano* 'across':

- (MTA.40)    Na    va    gap    iake    **vano**.  
               1s:AO    go    just    here    across  
               'I'm just going this way (level, parallel to shore).'

[FP1-41b]

This being said, Mota shares one property with Löyöp, namely a lexical split of the land-sea axis. In the higher parts of the island, the vertical directionals *up-down* are used; but on the lower areas, namely the coastal village and the sea, different terms are used. One notable difference with Löyöp, though, is the asymmetry of Mota directionals. Indeed, the lexical split concerns only the *seaward* direction, the one that runs from the island's top towards the ocean: it is encoded *swo* 'down' in the bush, and *rowo* 'out' elsewhere. Similarly, the directional used at sea, when pointing towards the ocean (#17), is never 'down' like in Gaua, but always *rowo* 'out'. As for the opposite direction, namely *inland*, it is consistently lexified *sage* 'up' regardless of the slope, or of the location on the island. The system of Mota is represented in Figure 14.

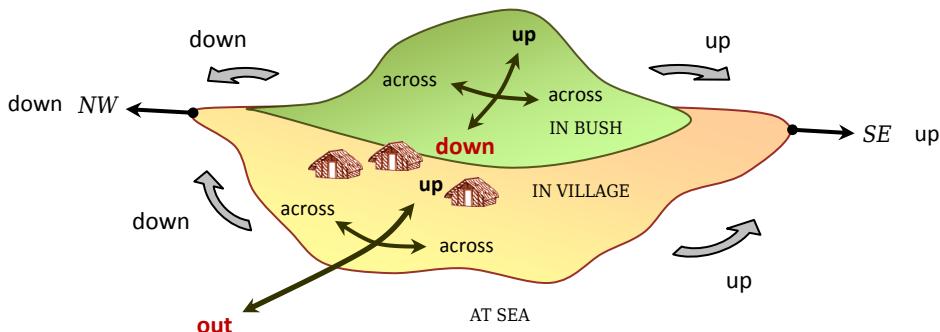


Figure 14 – The system of geocentric directionals in Mota

Keeping in mind the canonical analysis exposed in §4.4, one could say that Mota patterns everywhere like the canon Gaua, except for seawards directions on flat terrain (#16-17 vectors in Table 6 p.16), for which Mota follows the same strategy as the other canon Mwotlap. This asymmetry is reminiscent of the one we saw for Vanua Lava languages in §5.2.1; for the latter, it was a property of the coastal axis (Figure 12 p.38), whereas Mota is asymmetrical on the land-sea axis.

The following examples, based on the kinetic presentative *veta* (+directional),<sup>18</sup> illustrate the asymmetry of directionals:

(MTA.41) Nea ilunia veta **sage**.

3sg there PRSTV up

- a) <CARDINAL> 'There he is, walking towards southeast (along the shore).'
- b) <TOPOGRAPHIC> 'There he is, walking uphill (on a slope).'
- c) <TOPOGRAPHIC> 'There he is, walking inland (on flat terrain).'

[FP1-41b]

(MTA.41') Nea ilunia veta **swo**.

3sg there PRSTV down

- a) <CARDINAL> 'There he is, walking towards northwest (along the shore).'
- b) <TOPOGRAPHIC> 'There he is, walking seawards (on a slope).'

[FP1-41b]

(MTA.41'') Nea ilunia veta **rowo**.

3sg there PRSTV out

'There he is, walking seawards (on flat terrain).'

[FP1-41b]

Example (42), taken from a narrative, shows the two strategies in sequence. After they finished carving their wooden canoe in the bush, the characters brought it *down* (#15) all the way to the beach; at which point they took it *out* (#16-17) to the ocean:

(MTA.42) Rave sur o nati aka **swo**, me vega kalo  
pull downwards ART small canoe down PFT climb upwards

i vawo nati aka, neira me va **rowo** ilo lama.  
ALL top small canoe 3pl PFT go out ALL open.sea

'So they dragged their small canoe (*down*) to the shore, climbed upon it,  
and *out* they went into the ocean.'

[MTA.Snake.34]

One would expect that the directional 'out' should contrast with its antonym 'in', like in the symmetrical systems of Löyöp or Mwotlap. However, this is not what we find in Mota, where *sage* 'up' is used in all cases. Arguably, this asymmetry of geocentric directionals can simply be explained by the existence of the same asymmetry in the TOPOLOGICAL domain: as we saw in §2.4.2 (Table 5), Mota has innovated a lexical contrast between 'down' (*swo*) and 'out' (*rowo*), but has kept its direction 'in' colexified with 'up' (*sage*) - cf. (22) p.15.

---

<sup>18</sup> For a detailed description of the kinetic presentative in the neighbouring language Mwotlap, see François (2003:156-162).

### 6.3 A distance-based lexical split: Mwerlap

#### 6.3.1 Two nested subsystems

Mwerlap, the language spoken on Merelava island, can also be described as a hybrid between the two canons of Gaua and Mwotlap. Like Gaua, Mwerlap has preserved the undifferentiated traverse ('across') on the local scale, and sometimes encodes the land-sea axis as *up-down* (**seag-sōw**); however it shares with Mwotlap, at least in some cases, the lexification of that same land-sea axis as *in-out* (**sar-row**).

Based on these preliminary observations, one could propose to see Mwerlap simply as a variant of Löyöp. But such an analogy would fail to take into account an ingredient specific to Mwerlap: namely, that the variation is not based on the slope, but on physical distance. Remember that in Löyöp, the main criterion for choosing between *up-down* and *in-out* on the land-sea axis, was the declivity of the ground: the former pair was used when verticality was salient, the latter when the direction was roughly horizontal—whatever the distance. In the case of Mwerlap, the explanations given by my consultants mentioned distance as a critical factor in encoding the land-sea axis. According to them, the pair of directionals *in-out* must be used for very local reference, within a radius of about 20 metres on each side; whereas *up-down* remain the relevant directionals for farther distances. The resulting system is represented in Figure 15.

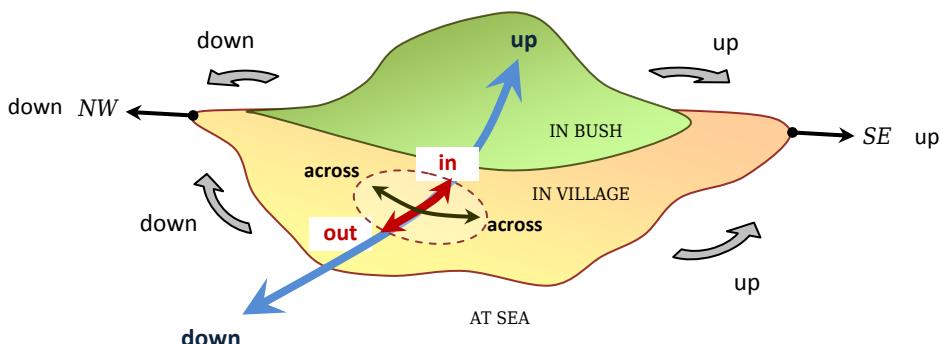


Figure 15 – The two nested subsystems of geocentric directionals in Mwerlap

Mwerlap thus involves two subsystems that appear to be nested. On the one hand, LONG DISTANCES employ two distinct *up-down* axes (both conservative, and used in Gaua languages); on the other hand, SHORT DISTANCES in the local setting resort to a different set of axes: *in-out* for the land-sea axis, and *across* for the traverse that crosscuts it. In terms of actual distances, estimations by my consultants mentioned a radius of about 20 m for the array of *in-out*, but a longer array of perhaps 200 m each side for the use of *across*. In other words, the “local” subsystem is less of a circle than an ellipse or strip parallel to the shoreline.

Sentences (43-44) in Table 9 exemplify the spatial configuration of Mwerlap. Such utterances can be heard during locally popular ball games such as soccer or volleyball (cf. *pas*, a loan from Eng. *pass the ball*). The forms of the directionals themselves will be explained in the next subsection; at this point, what matters is to illustrate the idiosyncratic organisation of Mwerlap geocentric directionals on the land-sea axis.

Table 9 — The distance-based lexical split of Mwerlap directionals on the land-sea axis

SHORTER DISTANCES (<20 m)		LONGER DISTANCES (>20 m)	
(MRL.43a) Pas <b>ser=lēg</b> ! pass <b>in</b> =thither 'Pass (him) the ball inland!' [BP3-21a]		(MRL.43b) Pas <b>sege=lēg</b> ! pass <b>up</b> =thither 'Pass (him) the ball inland!'	
(MRL.44a) Pas <b>ru=mē</b> ! pass <b>out</b> =hither 'Pass (me) the ball seawards!'		(MRL.44b) Pas <b>su=mē</b> ! pass <b>down</b> =hither 'Pass (me) the ball seawards!'	

The impact of physical distances upon the choice of directional is obvious: as the distance becomes greater than about 20 meters, the pair of local directionals *in-out* has to give way to the *up-down* directionals that belong to the wider perspective. Despite its differences with Löyöp and Mota (compare Figure 15 with Figures 13 and 14), overall the system of Mwerlap can also be said to ultimately revolve around declivity, in its own particular way. Indeed, the island of Merelava is a very steep, conical volcano, so relatively flat areas will always be narrow strips of gentle slope within a general shape of strong declivity.<sup>19</sup> In this regard, the DISTANCE-based system of Mwerlap is not that different from a DECLIVITY-based system such as that of Löyöp.

The manner in which Mwerlap structures its directionals based on distance is not found anywhere else in the area. While physical distance commonly results in different terms on the coastal axis (see §3.1), Mwerlap is the only language for which differences in distance also govern the choice of directionals on the land-sea axis.

### 6.3.2 Morphology of directionals in Mwerlap

Another intricacy of Mwerlap is the actual form taken by the directionals themselves. Most languages of northern Vanuatu have a single set of directionals, usually monosyllables, which are all formally invariant; these are the forms given in Table 6 p.16. By contrast, Mwerlap directionals vary morphologically depending on whether they define a static location or a motion path (cf. §2.3), and also on their combination with deictics. I give all the forms in Table 10, and explain them below.

Table 10 — Morphology of directionals in Mwerlap

	<i>topological meaning</i>	<i>directional + deictic</i>	Kinetic use		Static use	
			EGOCENTRIC	ALLOCENTRIC	EGOCENTRIC	ALLOCENTRIC
'inland' (<20 m)	'in'	ki <b>ser</b> ( <i>kē</i> )	<b>ser=mē</b>	<b>ser=lēg</b>	kere= <b>ser=mē</b>	ka(ra)= <b>sar</b>
'seawards' (<20 m)	'out'	<b>kor</b> ( <i>kē</i> )	<b>ru=mē</b>	<b>ru=lēg</b>	kere= <b>ru=mē</b>	ka(ra)= <b>row</b>
'// shore' (<200 m)	'across'	ki <b>vel</b> ( <i>kē</i> )	mē	<b>vel=lēg</b>	kere= mē	ka(ra)= <b>van</b>
'inland' (>20 m)	'up'	ki <b>sea(g)</b> ( <i>ñē</i> )	<b>sea(g)=mē</b>	<b>sege=lēg</b>	kere= <b>sea(g)=mē</b>	ka(ra)= <b>seag</b>
'seawards' (>20 m)	'down'	<b>kos</b> ( <i>ñē</i> )	<b>su=mē</b>	<b>su=lēg</b>	kere= <b>su=mē</b>	ka(ra)= <b>sōw</b>
'to SE' (>200 m)	'up'	—	<b>sea(g)=mē</b>	<b>sege=lēg</b>	<b>sea(g)=mē</b>	<b>seag</b>
'to NW' (>200 m)	'down'	—	<b>su=mē</b>	<b>su=lēg</b>	<b>su=mē</b>	<b>sōw</b>

<sup>19</sup> The colony of Mwerlap speakers who have established themselves on the east coast of Gaua island use the same system as on their home island of Merelava.

### ❖ STATIC LOCATIONS

The forms given in Table 6 for Mwerlap (namely **sar** ‘in’, **row** ‘out’, **seag** ‘up’, **sōw** ‘down’, **van** ‘across’) are the same as the rightmost column of Table 10; however, the basic forms are hardly used alone, and usually combine with other particles.

When preceded by the particle **ka** or **kara**, the directional defines a vector pointing to a static location, and deictically oriented away from the speaker (“allocentric”) — e.g. **ka(ra) seag** ‘up there’, **ka(ra) van** ‘over there [parallel to shore]’, etc.

- (MRL.45) Gil    **kara**    **row**!  
             dig    STATIC    out  
             (*digging a hole*) ‘Dig further away [from me], towards the sea.’

[BP3-20b]

Conversely, when the vector defined by the static location is deictically oriented towards the speaker (Eng. ‘up here, up this way’), the phrase is followed by the enclitic **=mē** ‘hither’. This use is not problematic *per se*, and simply corresponds to the “egocentric” use of the deictic directional ‘hither’ that was presented in §2.4.1 above (cf. Table 4 p.12). But the peculiarity of Mwerlap is that this clitic **=mē** triggers leftward VOWEL HARMONY upon its host phrase, resulting in allomorphic forms of the directionals with raised vowels. Thus **kara sar** [kara'sar] ‘over there inland’ becomes **kere ser=mē** [kere'ser'mi] ‘over here inland’; **kara row** [kara'rōw] ‘over there seawards’ becomes **kere ru=mē** [kere'rū'mi], etc.

- (MRL.45') Gil    **kere**    **ru**    **=mē**!  
             dig    STATIC    out    hither  
             ‘Dig a bit more this way, towards the sea.’

[BP3-20b]

Because the directional ‘across, parallel to shore’ is lexified with **van** ‘thither’ which is originally allocentric (§2.4.1), its egocentric counterpart is not **\*kere ven=mē**, but simply **kere mē** ‘this way’.

### ❖ MOTION PATHS

When the directional vector defines a motion path followed by a participant, the directionals combine with the enclitic **=mē** for egocentric orientation (**su=mē** ‘down this way’) and **=lēg** when allocentric (**su=lēg** ‘down that way’).<sup>20</sup> This kinetic use of directionals was illustrated in sentences (43-44) above, in which the motion path outlined by the directionals was the one followed by the ball.

### ❖ COMBINATION WITH DEICTICS

Finally, Mwerlap directionals shows special forms when they are combined with a demonstrative. The rich system of Mwerlap demonstratives include **kē** ‘PROXIMAL’ (with variants **kēkē**, **kēlē**...) as opposed to **ñē** ‘DISTAL’ (with variants **ñēñē**, **ñea** ...). Vowel harmony in the locative phrase sometimes triggers the raising of the directional’s vowel (e.g. **ki ser kē** ‘inland here’). As Table 10 above shows, some forms are unpredictable, such as {**ki+row=**} **kor** ‘out’ and {**ki+sōw=**} **kos** ‘down’.

<sup>20</sup> The forms for ‘up’ are unpredictable, respectively **sea=mē** and **sege=lēg**. The ones for ‘across, parallel to shore’ are **vel=lēg** (‘thither’) if allocentric, and simply **mē** (‘hither’) if egocentric.

The following examples illustrate the directionals when they are combined with a demonstrative.

- (MRL.46) i Edga **ki** **vel** **kēlē**.  
 PERS Edgar LOC across PROX  
 'Edgar is over there this way (parallel to shore).' [BP3-20b]
- (MRL.47) i Edga **kor** **kē** verē.  
 PERS Edgar LOC:out PROX outside  
 'Edgar is (out) here outside.' [BP3-20b]
- (MRL.48) Sean mē-lē **sar** lē eañ **kos** nē.  
 3sg PFT-take in LOC house LOC:down DIST  
 'He took it [the knife] into that house *down* over there (seawards).' [MRL.d05:20]
- (MRL.49) Ne-tedun irō se-velvelēlē **vel** nē lē sar.  
 ART-person two IPFV-argue across DIST LOC clearing  
 'Two people are arguing over there in the middle of the village.' [MRL.d08:02]

The complex system of space reference in Mwerlap would certainly warrant a study in its own right. However, the information given here should serve our main purpose here, namely the cross-linguistic comparison of geocentric directional systems.

## 7 Lexical innovation and resulting asymmetries: Lehali, Lo-Toga

Until now, all the systems examined so far were associated either with a single language, or with a set of languages spoken on a single island. Indeed, it makes sense for communities living on the same island – such as Gaua, or Vanua Lava – and involved in regular contact, to share the same system of geocentric reference. This is not an absolute rule, though, as we will see now.

The system to be described in this section is shared between Lehali and Lo-Toga, which makes it unusual in two respects. First, it means that Lehali uses a directional system distinct from Löyöp, although both languages are spoken on the same island of Ureparapara.<sup>21</sup> This may be explained, historically, by the fact that the modern speakers of Löyöp descend from a community that was initially based in the neighbouring atoll group of Roua (François 2012: 97), and only came to Ureparapara during the 20th century. The other originality of the system here described is that it is used across different islands: the eastern coast of Ureparapara island (northern Banks) for Lehali, and the three southern islands of the Torres group (Toga, Lo, Tegua) for Lo-Toga. This alone shows that a given system is not necessarily tied to an island's specific topography, and can diffuse from language to language through language contact.

A corollary of that situation is that there is no short way to refer to this system. Whereas other systems so far could be designated using the name of the language ("the Volow system") or of the island it was linked to ("the Gaua system"), the one that is shared between Lehali and Lo-Toga will simply be called here the "LLT system".

---

<sup>21</sup> To be fair, we already saw that Mwotlap and Volow show slightly different configurations, even though they are spoken on the same island of Motalava.

## 7.1 Partial reflexification of ‘up’ directional

The LLT system of geocentric directionals can be seen as a variant of Löyöp which was presented in Figure 13 p.39. Like Löyöp (and Mwotlap), the cardinal axis has been generalised everywhere on land, for directions parallel to the coast. Like Löyöp, the land-sea axis shows a lexical split between two strategies depending on the location on the island: *in-out* directionals are used on lower areas of the island, and *up-down* in the steeper parts.

However, compared with Löyöp, the LLT system has a peculiarity: it shows two different directionals for ‘up’. One of them is *ag* [au] or *iag* [iau] in Lo-Toga, and *ha* [ha] in Lehali: these are regular reflexes of the Proto Oceanic etymon for ‘up’, namely *\*sake* (cf. Figure 5 p.19), which was also the origin of all ‘up’ directionals seen so far in other Banks languages – e.g. Volow *ha* [ha], Lakon *hag* [hau]. The other directional is *vin* [βin] in Lo-Toga, *vēn* [βm] in Lehali: these reflect a recent innovation from a local verb *\*vene* ‘climb, go up’. In sum, Lehali and Lo-Toga have an inherited directional ‘up’ (<*\*sake*) and an innovative one (*\*vene*), when all other languages seen so far had only one form.

Crucially, these two directionals present different distributions in the modern geocentric system. The innovative ‘up’ directional (*\*vene*) has replaced the older one for the vertical meaning ‘up’ – as in ‘up in the tree’ – and it is the directional used when pointing inland in the steeper parts of the island, i.e. that area where verticality is most salient. By contrast, the older ‘up’ directional *\*sake* has been restricted to the cardinal axis, whether it serves for long-distance navigation across islands (§3.3), or for the encoding of directions parallel to the coast, pointing southeast (see Maps 3 and 4 p.23).

## 7.2 A strictly synchronical view

In fact, if we were to take a strictly synchronical approach, we shouldn’t be speaking of two ‘up’ directionals. Insofar as the innovative *\*vene* directional is the only one that can translate ‘up’ on the vertical axis, it is the only one that can be legitimately glossed ‘up’ in a synchronic representation of the modern LLT system. As for the other term *\*sake* – historically ‘up’ > ‘upwind’ > ‘southeast’ – it doesn’t mean ‘up’ any more: its only meaning in modern Lehali and Lo-Toga is that of a geocentric directional. Lo-Toga (*iag*) and Lehali *ha* have now lost any semantic connection with the vertical axis, and are only ever used on the horizontal plane, as a cardinal term pointing southeast.<sup>22</sup>

The modern system of Lo-Toga and Lehali is represented in Figure 16. I here gloss ‘up’ only that directional which can also take, in synchrony, a vertical meaning; the other directional (*\*sake*) is glossed ‘southeast’, its only modern meaning.

---

<sup>22</sup> Recall that this gloss ‘southeast’ refers to what can be reconstructed as the underlying principal orientation associated with the cardinal axis; however, when used on land, the actual reference of the vector may undergo a certain degree of rotation depending on the shape of the shoreline (see §3.4.1).

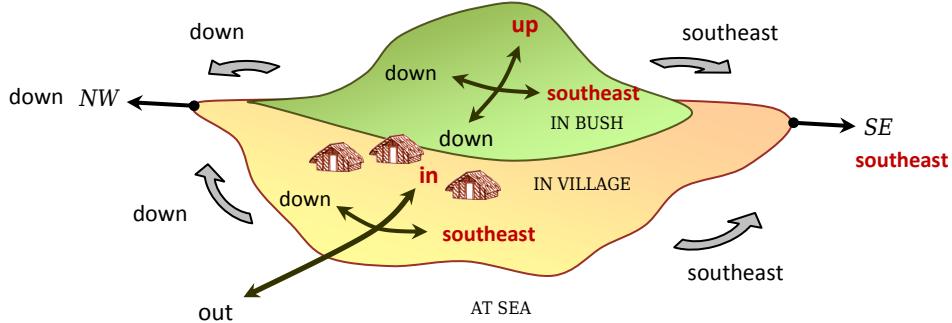


Figure 16 – The system of geocentric directionals in Lehali and Lo-Toga

### 7.3 Examples

The following Lo-Toga examples illustrate the LLT system. Pointing ‘inland’ from a location on the sea, or on the flatter parts of the island, involves the directional ***il*** ‘in’. When it has its geocentric meaning, the directional is often better translated using adverbial locative phrases in English – such as ‘on the beach’, ‘ashore’, etc. – or not translated at all:

- (LTG.50) Pahwëne nihe ge= vēn ***il*** ***me***, Merawehih v=***il*** hag.  
then 3pl AO:pl= go **in** hither (hero) IPFV=**in** sit  
Ni= itë nihe ge= rōw ***il*** ***me***, nie ni= vē ***rōw*** ***me***  
AO:3s= see 3pl AO:pl= rush **in** hither 3sg AO:3s= go **out** hither  
ni= ere teletale n=ēke.  
AO:3s= smash:SG in.pieces ART=boat

‘As they were getting closer to the island [liter. coming **in**], Merawehih was waiting on the beach. As soon as he saw them landing *ashore*, he walked *down to them* [liter. he came **out**], and suddenly smashed their boat into pieces.’ [LTG.Merawehih.053]

But the same INLAND direction is encoded with ***vin*** ‘up’ if it points to the bushy areas of the island:

- (LTG.51) Ne=lete mi kemor na in revtë, vet ne ***vin*** in.  
ART=garden POSS 1ex:du STAT lie close place REL **up** lie  
‘Our garden is very close, this way *up(hill)*.’ [LTG.d01:15]

In this case, the term pointing seawards is not ***rōw*** (*out*) any more, but ***iw*** ‘down’:

- (LTG.52) Kemë ve=toge deh=***vin***, pa heqere ha ve=toge dih=***iw***.  
1ex:pl IPFV=stay side=**up** but HUM:PL other IPFV=stay side=**down**  
‘We live *uphill*, but there are other people who live *downhill*.’ [FP1-13a]

The latter use of ***iw*** ‘down’ on the land-sea axis must be distinguished from its use on the cardinal axis. In the latter case, ***iw***<sup>23</sup> contrasts not with the directional ***vin*** ‘up’, but with the directional ***(i)ag*** ‘southeast’.<sup>24</sup>

- (LTG.53) N=ēñwe mēhe ve=tu vet ne v=***ag*** in,  
ART=house their IPFV=stand place REL IPFV=southeast lie

<sup>23</sup> In (53), the directional ***iw*** presents its allomorph ***w'***, which is regular when followed by a vowel.

<sup>24</sup> See also ex.(23) p.21 for another example of ***(i)ag*** ‘southeast’ on the navigational scale.

si    vet    ne    **w'**    in?  
 or    place    REL    down    lie  
 'Is their house located on the southeast side, or the northwest side?'

[FP1-13a]

In sum, when the directional ***iw*** 'down' means 'downhill' (land-sea axis) it contrasts with the historically innovative term ***vin*** 'up'; but when it takes its cardinal meaning 'northwest < downwind', it contrasts with the vestigial form (***i***)***ag***, whose meaning in synchrony is purely geocentric. The same situation, *mutatis mutandis*, prevails for the language Lehali.

#### 7.4 Total vs partial relexification

The complex situation of Lehali and Lo-Toga results from a process of PARTIAL RELEXIFICATION.

In previous sections, we had already come across cases of relexification, as one inherited word had been replaced historically with an innovative one.<sup>25</sup> However, most cases involved TOTAL RELEXIFICATION, in the sense that a term A was replaced by a term B for the totality of its uses. Consider, for example, the case of Proto Oceanic \**sipo* 'down': in four languages of Gaua, this etymon was replaced by another lexeme \**roro*, whose original meaning can be reconstructed as 'go deep, sink; be deep, be low' (François 2010:139): this is the source of the modern forms *ror* [rɔr] in Nume, Dorig, Koro, and of *roy* [roj] in Olrat. Crucially, \**roro* became the standard directional for 'down', and replaced \**sipo* for all of its uses, whether on the vertical axis or on the horizontal plane, whether on the land-sea or on the cardinal axis. The semantic distribution of *hōw* in Lakon and *roy* in its neighbour Olrat are identical (Table 6 p.16), in such a way that the semantic systems remain isomorphic in spite of the lexical replacement (see François 2010:139, 2011:226-228). The same demonstration could be made with Lo-Toga *il*, Lehali *ila* ('in, inland'), whose semantic outline matches exactly that of Löyöp *say* ('in, inland') despite the lack of cognacy. Such processes of total relexification do not affect the semantic organisation of the lexicon, and therefore have limited impact upon the way the system works.

The type of relexification that took place in Lehali and Lo-Toga for UP directionals is of a different nature, because it was only partial. Among the various meanings that used to be associated with \**sake* 'up' (cf. Löyöp *sa*), only a subset were affected by the lexical change that replaced \**sake* with \**vene* – namely, the vertical sense and the geocentric sense most obviously associated with verticality, namely 'inland' in the steeper areas of the island. The result is a historical split between various senses that used to be colexified, and an effect of asymmetry, as the same directional 'down' ends up contrasting with two distinct directionals – as in (52)-(53) above.

The last system to be examined, namely Hiw, presents another example of asymmetry resulting from partial lexical replacement – yet an even more complicated one.

---

<sup>25</sup> For a general presentation of processes of lexical replacement in the Torres and Banks languages, see François (2011:203-210).

## 8 The puzzle of Hiw

### 8.1 A quirky system

The most intricate of all geocentric systems found in northern Vanuatu is no doubt the one used on Hiw, the northernmost island of the archipelago. Even though the language of Hiw is often close to its neighbour Lo-Toga, it also shows numerous idiosyncrasies that set it aside as a unique language. The geocentric system of directionals is no exception to this.

The Hiw system stands out, to begin with, if one considers its organisation in Table 6 p.16. Hiw is the only language of northern Vanuatu that possesses a directional ‘in’ yet does not provide it with any geocentric meaning. It is the only language in which the cardinal directional for ‘southeast’ used on land for long-distances (#9 *vēn*, literally ‘thither’) differs from the one used across islands (*ag* #8). Besides, the latter directional *ag* shows an odd pattern of colexification between #8 ‘southeast (across islands)’ and #4-5 ‘inland (on flat terrain)’ which is found nowhere else, and can hardly be given a simple gloss. All these oddities constitute puzzles that need to be solved.

The geocentric system of Hiw is represented in Figure 17, reproduced from Figure 3 p.5.

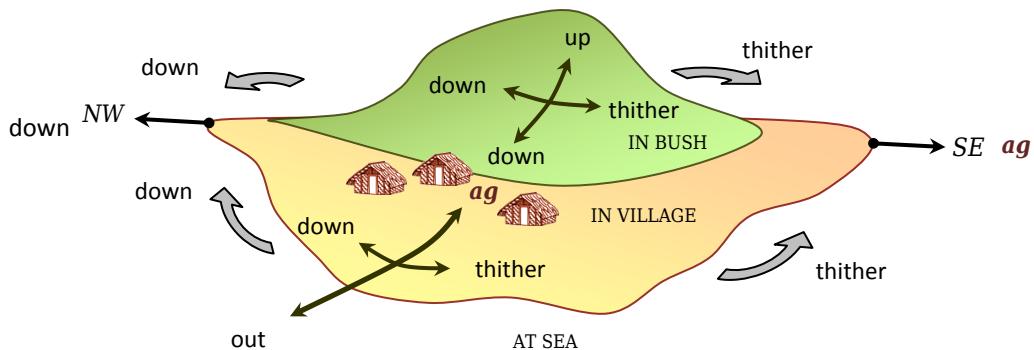


Figure 17 – The system of geocentric directionals in Hiw

The impression of quirkiness left by the Hiw system depends, of course, on the point of comparison. For example, Hiw has very little in common with the languages of Gaua (§1.2, 4.2): except perhaps for the use of DOWN for ‘northwest’ and ‘downhill’, everything else is different. The difference becomes less extreme if Hiw is compared with its immediate neighbours, such as Lo-Toga or Löyöp. In the remainder of this study, I will argue that the system of Hiw, aberrant as it is in synchrony, can be accounted for by reconstructing a number of innovations – most of which also took place in other northern Vanuatu languages.

### 8.2 An unusual colexification between ‘thither’ and ‘southeast’

#### 8.2.1 /Vēn/ as a participant-oriented directional

Figure 17 assigns to vectors #9-10 a gloss ‘thither’. This is justified by the principle underlying these representations (§1.2), which consists in glossing a geocentric directional with the non-geocentric meaning it has in the same language, considered in synchrony.

And indeed, *vēn* exists in Hiw as a participant-oriented directional (§2.4.1; row #2 in Table 6). Its meaning is allocentric, i.e. it is semantically directed at a participant outside the speaker’s sphere, and typically translates ‘to you/him/them...’. In (54), *vēn* gives the

instruction to retrieve a participant from the context, namely the mother. Had nobody been in the pit, the speaker would have resorted to a non-personal strategy such as *uw* ‘down’.<sup>26</sup>

- (HIW.54) Sōrō giy ne=qeřoň tēn, giy tēvěkqö, ūe= suř i ūakña-se  
 3du dig ART=hole ground dig deep AO:du= install DOM mother-3NSG  
*uw* yōne. Viye nōn ūakña-se yite, ne= gengon,  
**down** inside take:PL POSS mother-3NSG firewood ART= food  
 viye **vēn** eyō qeřoň tēn.  
 take:PL **thither** LOC hole ground

‘The [brothers] dug a pit in the ground, a deep pit, and installed their mother [down] inside. They gathered firewood for her as well as food, and brought it all *to her* [LITER. took it *thither*] in the pit.’

[HIW.Brothers.07]

This deictic use of *vēn*, glossed ‘thither’ for a shortcut, contrasts with *me* ‘hither’:

- (HIW.55) Viye qē **me** ti noke! Noke viye **vēn** ti ne= sōgē =kye on gon.  
 take:PL SUGG hither DAT 1sg 1sg take:PL thither DAT ART= pig =my SBJN eat  
 ‘Give them to me! I’ll give them to my pig for food.’

[HIW.d09:41]

Despite their phonetic similarity, *vēn* [βen] ‘thither’ and *vēn* [βm] ‘up’ are two distinct directionals. Both are found in (56), where a motion is first described in geometric terms (‘climb up’) and then explicitly anchored to a participant (‘climb towards him’):

- (HIW.56) Nine yō ne=megoye kkē in **vēn** sag. Nine vēn **vēn**, vēn wate **vēn**.  
 3sg see ART=child small DEF **up** sit 3sg climb **up** climb reach **thither**  
 ‘He saw the small boy sitting *up* (in the tree).  
 So he climbed *up*, he climbed all the way (to *him*).’

[HIW.Music.19]

### 8.2.2 /Vēn/ on the coastal axis

The participant-oriented use of *vēn* may correspond to any vector in spatial terms. In addition, the same directional *vēn* is also found with a geocentric meaning, in which case it constructs a vector parallel to the coast, specifically oriented southeast (#9-10 in Figure 4 p.17). Obviously, its actual orientation on a given point of the island may correspond to a variety of surface directions – as delineated by the arrows in Map 2 p.23 – but its underlying orientation can be reconstructed as ‘southeast’, following the reasonings of §3.4.3.

- (HIW.57) N=ēñwe =ma owēte **vēn** taqe.  
 ART=house =1ex:pl PRSTV (southeast) stoop  
 ‘Our house is over there *this way* (southeast).’
- [EP2-17b]
- (HIW.58) Sōrō tō **vēn** ñwē ne, ūe= tō wōywōy vaviyi **vēn**.  
 3du go:NPL up like this AO:du= go:NPL crosswise side (southeast)  
 ‘They walked uphill like this, and then veered towards southeast.’

[FG2-14b]

In this geocentric sense, *vēn* contrasts with *uw* ‘down; northwest’:

- (HIW.59) Ike tati sesō **uw!** Ike sö **vēn** ti ne= Yugemëne.  
 2sg NEG paddle:RED down 2sg paddle (southeast) DAT ART= (village)  
 ‘Don’t paddle north! You should paddle southeast, towards Yugemëne.’

[FG2-14b]

<sup>26</sup> See also the pair of sentences (12)-(12') above.

Note that *vēn* is used for any southeast vector on the local scale, whether in the village (57), in the heights of the island (58) or on the sea along the coast (59).

Considering the contrast with *uw* ‘down’, and the observation that Southeast is encoded as ‘up’ in all other northern Vanuatu languages, the Hiw strategy is puzzling. The phonetic closeness of this [βen] with the vertical ‘up’ [βm], which some younger speakers initially described as mere homophones, confused the picture even more in the earlier phases of my exploration. However, elder speakers confirmed that the directional used for southeast on land was distinct from ‘up’, and instead homophonous with ‘thither’.

### **8.2.3 A memory left by the traverse axis**

How can this colexification ‘thither’/‘southeast’ be explained? Incidentally, a similar situation is found with Lemerig *wēl*, which colexifies ‘across (towards SE)’ and ‘thither, towards non-speaker’. In the case of Lemerig, the etymology of *wēl*, from \**polo* ‘across, crosswise’ (§5.2.2) suggests that its use on the traverse axis (‘across’) must have been its primary meaning. As for the deictic use ‘thither, towards non-speaker’, it appears to be a secondary extension of that primary non-deictic meaning.

The situation in Hiw is similar to Lemerig, yet slightly different. My hypothesis is that the two directionals *vēn* form an etymological doublet rooted in the original polysemy of its etymon \**pano*. Indeed, *vēn* is the regular reflex of Proto Oceanic \**pano*, which can be reconstructed as a directional verb ‘go away; move in a transverse direction’ (Ross 2007: 279). It is likely that Proto Oceanic used \**pano* on the navigational scale (cf. Figure 5 p.19) but also on the local scale, this time encoding the undifferentiated traverse (§4.2.2). The ancestor of Hiw therefore used \**pano* to lexify both vectors #10 and #11, just as it still does in Mota or Nume.

Later on, the generalisation of the cardinal axis to all scales on land, which took place completely in Mwotlap (§4.3) or Löyöp (§6.1), for some reason was only complete in Hiw on the ‘downwind’ direction, without affecting the ‘upwind’ side. What resulted was an asymmetrical configuration: ‘down’ is used on the northwest side whatever the distance, whereas the southeast side maintains a distance-based split between *vēn* for shorter distances, and *ag* (<from \**sake*) on the navigational scale. Interestingly, the same asymmetry was found in the languages of Vanua Lava (§5.2): a single directional on the NW side, contrasting with two directionals for SE directions. In both cases, the SE directional for shorter distances is a vestigial trace left by the earlier transverse axis used on the local scale.

A final tweak added by Hiw was the generalisation of *vēn* to all vectors on land, including on the “intermediate scale” of long distances on a single island. As a result, Hiw is the only northern Vanuatu language that encodes southeast for long distances on land (*vēn* <\**pano*) differently from southeast across islands (*ag* <\**sake*): compare the rows #8 and 9 in Table 6 p.16.

### **8.2.4 An etymological doublet**

The second part of the scenario is the observation that while POC \**pano* evidently had the meaning ‘move in a transverse direction’, it can also be reconstructed with a sense ‘leave, go away (from speaker)’. The deictic component of this meaning (‘away from speaker’) explains why the six northernmost languages of the Torres-Banks area have grammaticalised it into a

participant-oriented directional ‘towards non-speaker’ – as shown in Table 3 p.9. This is the source of the allocentric directional in Mwotlap (*van*), Lo-Toga (*vēn*) but also Hiw (*vēn*).

For the same meaning, the nine southernmost languages of the Banks, from Vurës down to Mwerlap (Table 3) reflect another deictic directional verb of POc, namely \**watu* ‘go towards addressee’ (Ross 2007:275) – e.g. Lakon *at*, Vurës *net*, etc.; in doing so, they broadened its semantic scope not just to the addressee (\**watu* ‘towards you’) but to any participant outside the speaker’s sphere (‘towards you/him/her/them/it’...).

In sum, the two modern senses of Hiw *vēn* originate in the polysemy of its etymon \**pano*. This situation is summarised in Table 11. Glosses for POc reconstructions come from Ross (2007). I compare Hiw with two other languages, Mota and Mwotlap.

Table 11 – Explaining the homophony of ‘thither’ and ‘southeast along the coast’ in Hiw

POc etymon	Mota	Hiw	Mwotlap	DIRECTIONAL GLOSS
* <i>pano</i> ‘move in transverse direction’	 <b>vano</b>	 <b>vēn</b>	(CARDINAL)	‘across, along the coast’ ‘along the coast, SE side’
* <i>pano</i> ‘go away (from speaker)’	 <b>at</b>	 <b>vēn</b>	<b>van</b>	‘thither, to non-speaker’
* <i>watu</i> ‘go to addressee’				

Hiw is the only language that has kept reflexes of \**pano* both for the coastal axis and for the allocentric directional. The first of these two uses is also found in Mota, and is conservative of POc; it has been lost in a language like Mwotlap, which has replaced it with its cardinal directionals *up-down*. As for the deictic use ‘thither’, it is only found in northern Vanuatu languages, but not in languages like Mota. Hiw is partly conservative like Mota, and partly innovative like Mwotlap: this explains the presence of this etymological doublet in the Hiw system of directionals.

### 8.3 Another lexical split of UP

#### 8.3.1 Two homophonous directionals

The second puzzle of Hiw is its directional *ag*. I didn’t provide it with a gloss in Figure 17 above, because the term is only ever used with a geocentric meaning, and is found nowhere else in the language. Admittedly, the same could be said of the homophonous directional *ag* in the language Lo-Toga (or *ha* in Lehali, for that matter); yet the latter was provided with a specific gloss ‘southeast’ in Figure 16, because its synchronical semantics were clear enough.

The problem with Hiw *ag*, is that it can hardly be given any consistent glossing, because it encodes two directions which have nothing in common: on the one hand, *ag* encodes ‘southeast’ for long-distance navigation across islands (vector #8 in Figure 4 p.17); on the other hand, it is the directional used on the land-sea axis on flat terrain, i.e. #4 ‘landwards’ at sea, or #5 ‘inland’ in a village. There is no reason why these two directions should be merged, as their underlying definition is quite distinct, and they seldom align. From a strictly synchronical point of view, the only reasonable decision is to posit two homophonous directionals: *ag*<sub>1</sub> ‘southeast on the navigational axis’ (#8), and *ag*<sub>2</sub> ‘inland, on flat terrain’ (#4-5).

The navigational *ag*<sub>1</sub> is illustrated in (60). Even though this *ag* differs in synchrony from the directional *vēn* ‘up’, it follows the logics of the ‘up(wind)’ directional described for other languages – see Figure 6 p.20:

- (HIW.60) Kema peon vēn vaviyi **ag** Gawe.  
1ex:pl FUT go side (navig:SE) Gaua  
'We'll be travelling southeast, to Gaua.'

[FG3-39b]

This *ag*<sub>1</sub> contrasts with *uw* ‘down > downwind, northwest’. For example, the native name of the Solomon islands in Hiw is *Take-siw-uw* ‘on the side that goes down’. And the name of Hiw island itself reflects POc \**sipo* ‘go down’, as it is located downwind with respect to all other islands of northern Vanuatu (§3.3).

The second directional *ag*<sub>2</sub> ‘inland’ is illustrated in (61).<sup>27</sup>

- (HIW.61) Vē—n vēn, se= vēn **ag** net-venyō kkē.  
go:DUR go 3pl:AO= go:PL (**inland**) DIMIN-island small  
Se= yē **ag**: ne=tayō ñot v= **ag** tu.  
3pl:AO= look (**inland**) ART=person INDF IPFV= (**inland**) stand  
"Tekñwa, pa yē v= **ag** tu ūrē?"  
people but who IPFV= (**inland**) stand DIST  
'After a while, they came closer (*landwards*) to a small islet.

They looked up (*inland*): someone was standing there (*on the shore*).

"Hey guys, who's that standing over there (*inland*)?"

[HIW.Meravtit.111]

The opposite of this *ag*<sub>2</sub> is *rōw* ‘out’:

- (HIW.62) Sise vēn se=rav ne=wake kkē =sa, on ñav wate **rōw** yō pēgone.  
3pl go:PL 3pl:AO=drag ART=boat small =their SBIN drag reach **out** LOC sea  
'They dragged their canoe all the way *down* [liter. *out*] to the sea.'

[HIW.Meravtit.051]

### 8.3.2 Another etymological doublet

In sum, a strictly synchronical approach to Hiw would have hard time explaining why the same directional is used for two very different geocentric functions. However, things become clear again once we take a historical perspective – helped with our earlier examination of other northern Vanuatu languages.

My hypothesis is that we are dealing again with an etymological doublet. Both *ag*<sub>1</sub> and *ag*<sub>2</sub>, phonetically [auŋ], are regular reflexes of POc \**sake* ‘go up’. Just like we saw for Lehali and Lo-Toga (§7), the semantic connection has been blurred by the fact that these three northern languages have lost \**sake* for the vertical direction, and replaced it with \**vene* ‘climb > upwards, up’. In the synchrony of Hiw, the directional ‘up’ *vēn* presents only one geocentric use: namely, the encoding of ‘inland, uphill’ in the bushy parts of the island:

- (HIW.63) Ne= sov =en ye **vēn** esos ūrē?  
ART= smoke =POSS who **up** IPFV~smoke DISTAL  
'Whose smoke is smoking over there *inland* [liter. *up*]?' [FP3-28b]

<sup>27</sup> The sentence is taken from the famous story of Meřavitit, the local name of Kpwet. It can be usefully compared with the equivalent excerpt which I recorded in Lo-Toga, (LTG.50) above.

In a way, we are faced here with the same sort of process that we already saw for Lehali and Lo-Toga: the inherited word for ‘up’, namely *\*sake*, has been replaced by an innovative form *\*vene*. What resulted was a lexical split. On the one hand, some of the vectors formerly associated with *\*sake* underwent replacement to *\*vene* – hence *vēn* for vertical ‘up’ + ‘uphill’ as in (63). On the other hand, other vectors remained lexified with reflexes of *\*sake*, but lost their semantic connection with vertical ‘up’. But while Lehali and Lo-Toga had only retained *\*sake* for a coherent set of vectors (all pointing ‘southeast’), the originality of Hiw is that it has preserved its directional *ag* (<*\*sake*) in two separate corners of its geocentric system.

The navigational *ag<sub>1</sub>* originates in the cardinal sense of *\*sake* ‘up > upwind > southeast’. As for *ag<sub>2</sub>*, its origin lies in the ancient connection of ‘inland’ with *\*sake* ‘up’, which is still preserved in southern Banks languages. While such a connection ‘up’/‘inland’ is to be expected in the context of Oceanic languages, it is nevertheless surprising in the local Torres-Banks context. Indeed, we have seen that all the ten northernmost languages of Vanuatu have consistently lost their directional ‘up’ when referring to ‘inland’ in a village context, and instead have all shifted to a contrast ‘in’–‘out’: this was visible in Table 6 (p.16), which shows that all languages that have a specific directional for ‘in’ (row #3) also use it for ‘inland’ at least in the lower areas of their island (row #5). The fact that Hiw uses *rōw* ‘out’ for the ‘seawards’ direction – as in (62) above – reinforces the expectation that the ‘inland’ vector should have been encoded as *\*iy* ‘in’ – just like in Hiw’s neighbour, Lo-Toga. Instead, what we observe is the unexpected retention of *\*sake* at a point in the system where it should have long disappeared.

Table 12 — Reflexes of POc *\*sake* ‘go up’ in a subset of northern Vanuatu languages

POc etymon	DIRECTIONAL GLOSS	MTA	MTP	LYP	LTC	HIW
<i>*sake</i> ‘go up’ → ‘upwind’	‘upwind = SE (inter-island)’	<b>sage</b>	<b>hag</b>	<b>sa</b>	<b>ag</b>	<b>ag<sub>1</sub></b>
	‘upwind = SE (on land)’	<b>sage</b>	<b>hag</b>	<b>sa</b>	<b>ag</b>	( <i>vēn</i> )
<i>*sake</i> ‘go up’ (vertical)	‘up (vertical)’	<b>sage</b>	<b>hag</b>	<b>sa</b>	( <i>vin</i> )	( <i>vēn</i> )
<i>*sake</i> ‘go up’ → ‘uphill’	‘uphill = inland (mountain)’	<b>sage</b>	( <i>hay</i> )	<b>sa</b>	( <i>vin</i> )	( <i>vēn</i> )
	‘uphill = inland (sea, village)’	<b>sage</b>	( <i>hay</i> )	( <i>hay</i> )	( <i>il</i> )	<b>ag<sub>2</sub></b>

In comparison with Lo-Toga, Hiw thus comes one step ahead in complexity. Not only did Hiw split the ‘up’ vectors in two blocks via lexical innovation (reflexes of *\*vene* vs reflexes of *\*sake*), but it also retained *\*sake* in two separate corners of its system. What results is another etymological doublet – on top of the one exposed in §8.2.4 – involving two vestigial directionals *ag*. The situation is summarised in Table 12, which displays modern reflexes of POc *\*sake* in a selection of five languages: Mota, Mwotlap, Löyöp, Lo-Toga and Hiw. Those forms which are not cognate with *\*sake* are given in brackets. The two traces left by *\*sake* in Hiw are semantically discontiguous, and constitute a doublet.

## 9 Historical synthesis: The accumulation of innovations

### 9.1 A compendium of Torres-Banks innovations

The complexity of the geocentric system in Hiw can be unravelled by reconstructing a number of successive innovations that transformed the original system of space reference in

Proto Oceanic. The fact that most of these innovations are also attested in other languages of the area confirms their likelihood for Hiw; incidentally, this also means that the special case of Hiw can be chosen as a convenient synthesis of the various changes that have affected space systems in the whole Torres-Banks region.

The initial point in the history of Hiw geocentric reference was Proto Oceanic, whose system was essentially preserved intact in the modern languages of Gaua. The following lines will therefore fulfil the promise of §1.2 in the introduction, that of providing a unified theory in order to account for the space system of Dorig (Figure 2 p.5) and that of Hiw (Figure 3), in spite of their stark differences.

The origin of all geocentric systems in Torres-Banks languages was evidently the one represented in Figure 18. The languages of Gaua (Figure 9 p.32) have preserved this system to this day; others have changed it to various degrees.

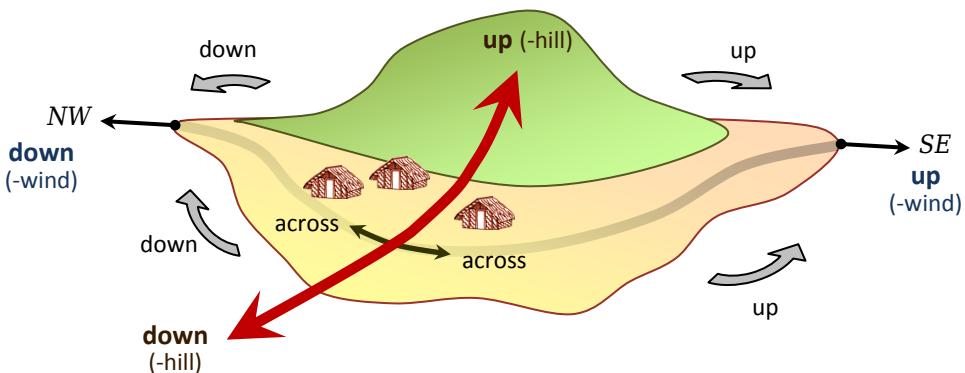


Figure 18 - The system of geocentric directionals in Proto Oceanic,  
the ancestor of all Torres-Banks languages (after François 2004)

### 9.1.1 Innovations on the land-sea axis

One widespread innovation has been the emergence of a new pair of topological directionals coding for 'in' and 'out' – two senses which Proto Oceanic (as well as the conservative languages of Gaua) originally colexified with 'up' and 'down', respectively (§2.4.2).

In all Torres-Banks languages outside Gaua, this new contrast *in-out* began to compete with the inherited pair *up-down* for the encoding of the land-sea axis. In Mwotlap and Vanua Lava, this innovation resulted in the complete redesign of the system, and the replacement of *up-down* by *in-out* everywhere for the land-sea axis. In other languages, though, the replacement was only partial: several languages preserved the *up-down* contrast in those parts of the island where verticality was cognitively salient (typically, in the bushy heights of volcanic islands) while they relexified the axis to *in-out* in the flatter parts of the island, or the land/sea interface. What resulted were hybrid systems, in which the land-sea axis is sometimes lexified *in-out*, and sometimes *up-down* (Mota, Mwerlap, the languages of Ureparapara and the Torres Is.).

Incomplete redesigns also resulted in the emergence of more idiosyncratic combinations, such as a contrast *up-out* in Mota. This configuration *up-out* was also found in Pre-Hiw (*ag* 'inland<up' vs *rōw* 'seawards<out'), with the particularity that the semantic connection of *ag* with 'up' was later lost when Hiw relexified its vertical 'up' with another word. This is how modern Hiw now lexifies its land-sea axis with two pairs of directionals: 'up' (innovative

form *vēn*) vs ‘down’ (*uw*) in the higher parts of the island; ‘inland’ (vestigial form *ag<sub>2</sub>*) vs ‘out’ (*rōw*) in the lower parts.

### 9.1.2 Innovations on the coastal axis

The innovations followed by Hiw on the coastal axis are also reminiscent of similar processes we observed in other languages.

The northern languages of the Torres-Banks area all gave in to the pressure of extending the use of the cardinal axis from its original navigational use for long distances, to the smallest distances on the local scale. Once again, this redesign was complete in a language like Mwotlap, but other languages show more complex configurations. To be precise, the ten northern languages agreed in generalising the cardinal directional ‘down’ (‘downwind, northwest’) to all scales, without exception; but for some unknown reason, several languages have maintained a distance-based split on the southeast side of that coastal axis. This is how Vanua Lava languages contrast a long-distance ‘up’ (<\**sake*) and a short-distance directional whose etymological meaning is ‘across’ (<\**volo*).

Hiw too shows a contrast between *vēn* ‘southeast (on land)’ – a trace left by POc \**pano* ‘go in a transverse direction’ – and *ag* ‘southeast (across islands)’, from \**sake* ‘up, upwind’. Due to the relexification of vertical ‘up’, the latter form *ag* has lost any connection with verticality. Like Lehalī *ha* and Lo-Toga (*i**ag*), it is now restricted to its use as a geocentric directional, whether as *ag<sub>1</sub>* ‘southeast (across islands)’ or as *ag<sub>2</sub>* ‘inland (on lower areas)’.

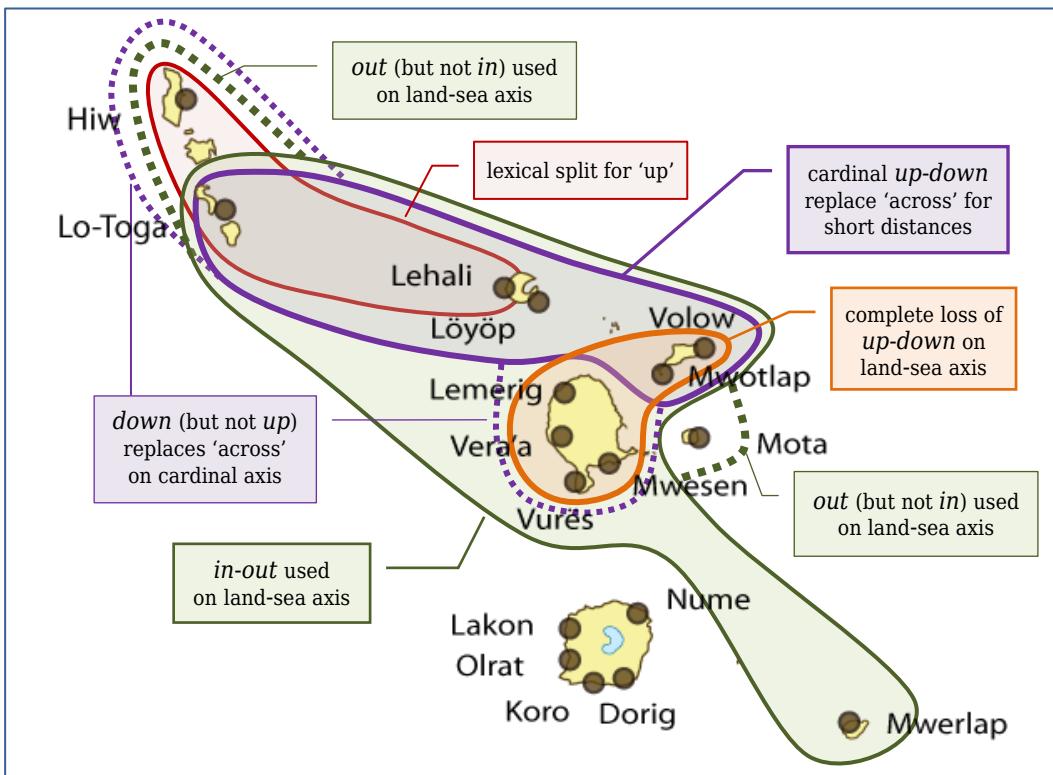
## 9.2 The perspective of linguistic geography

This study showed that the modern diversity of geocentric systems in northern Vanuatu results from a number of innovations which affected the original (Proto Oceanic) system, and accumulated during the course of history. While Table 8 above (p.34) reflected a synchronic approach, Table 13 organises the northern Vanuatu data following a historical perspective: it lists the six principal innovations involved in the make-up of modern systems, from the very conservative languages of Gaua (zero structural innovation from Proto Oceanic) to the more innovative languages in the north.

Table 13 — Main innovations involved in the development of modern geocentric systems in Torres-Banks languages.

	PARALLEL TO SHORELINE		ORTHOGONAL TO SHORELINE			
	DOWN replaces <i>across</i> (NW)	UP replaces <i>across</i> (SE)	IN used for ‘inland’	OUT used for ‘sewards’	total loss of UP/DOWN-hill	lexical split of UP
5 Gaua lgs	—	—	—	—	—	—
Mota	—	—	—	+	—	—
Mwerlap	—	—	+	+	—	—
4 Vanua Lava lgs	+	—	+	+	+	—
Mwotlap	+	+	+	+	+	—
Volow	+	+	+	+	+	—
Löyöp	+	+	+	+	—	—
Lehalī, Lo-Toga	+	+	+	+	—	+
Hiw	+	—	—	+	—	+

Evidently, these innovations were shared across neighbouring languages. The resulting isoglosses are represented in Map 10.<sup>28</sup> Interestingly, these isoglosses commonly intersect – a common situation in the northern Vanuatu linkage (François 2011; Kalyan & François, f/c). This can only be explained by a process of diffusion, whereby speech patterns – in this case, the internal structure of paradigms of space directionals – spread from community to community, via social and linguistic contact. Sometimes, neighbouring languages mutually aligned their systems perfectly, whether they were spoken on the same island (Vanua Lava) or on different islands (Lehali and Lo-Toga). In other cases, the alignment was only partial, as only some aspects of one system were able to leak from one speech variety to the other.



Map 10 – The modern diversity of Torres-Banks geocentric systems results from the accumulation of post-POc innovations that diffused across neighbouring languages

In sum, the history of space systems in the archipelago follows a pattern similar to what can be observed in the historical phonology or morphology of these languages (François 2011:201). The modern linguistic fragmentation, which today takes the form of *divergence*, ultimately results from the layering of various innovations, each of which once diffused – via a process of *convergence* – to a certain portion of the social network. Each community shared its innovations sometimes with one neighbour, sometimes with the other, ultimately resulting in the language mosaic that prevails today.

<sup>28</sup> Dotted lines reflect those cases when a new structure was adopted only partly, resulting in asymmetry. For example, while the solid purple line refers to the adoption of both cardinal *up* and *down* on the local scale, the dotted purple lines set apart those languages (Vanua Lava, Hiw) which generalised *down* but not *up*.

## 10 Conclusion

Considered in each language separately, the mechanism of geocentric space reference appears stable, at least within the time span of a few generations. But the perspective changes radically with systematic cross-linguistic comparison, as these space strategies appear in their fluidity and internal dynamics.

The comparative approach proves particularly helpful when attempting to interpret languages whose synchronic description unveils asymmetries and paradoxes. Such systems ultimately appear for what they are: mere moments in a history of ever-changing configurations. These constant adjustments, in turn, constitute possible answers to various cognitive and communicative pressures, such as the need to adapt to new landscapes, the avoidance of confusion, the tendency for analogical levelling, the entrenched cross-linguistic diffusion of innovations. Sometimes, in spite of these innovations, modern languages exhibit vestigial memories of earlier systems, which have been preserved against the odds of history.

Besides their intrinsical interest for a typology of space strategies, these directional systems provide an excellent vantage point when it comes to unravelling the multiple layers whereby linguistic systems are constantly reinvented by the populations who use them.

---

## 11 APPENDICES

### 11.1 Abbreviations

#### 11.1.1 Languages

The abbreviations for language names appear on *Map 1*, and are repeated below.

DRG	Dorig	LMG	Lemerig	MTP	Mwotlap	VRA	Vera'a
HIW	Hiw	LTG	Lo-Toga	MSN	Mwesen	VRS	Vurës
KRO	Koro	LYP	Löyöp	NUM	Nume	POC	Proto Oceanic
LHI	Lehali	MRL	Mwerlap	OLR	Olrat	PTB	Proto
LKN	Lakon	MTA	Mota	VLW	Volow		Torres-Banks

#### 11.1.2 Interlinear glosses

Example sentences are glossed according to the Leipzig rules. More specific abbreviations are listed below.

ALL	allative case	IPFV	imperfective
AO	Aorist (≈narrative) aspect	IRR	irrealis
ART	article	LOC	locative case
DEF	definite	MED	medial demonstrative
DIMIN	diminutive	NEG	negation
DIST	distal demonstrative	NPL	non-plural
DOM	differential object marker (human object)	NSG	non-singular
DX	deictic	PERS	personal article
EXIST	existential predicate	POSS	possessive classifier or linker
FOC	focus marker	POT	potential
HUM	human article	PRF	perfect
INDF	indefinite	PROH	prohibitive
INSTR	instrumental	PROX	proximal demonstrative
		PRSTV	Presentative aspect

RED	reduplication	STAT	Stative aspect
REL	relativiser	SUGG	suggestive (polite)
SBJN	Subjunctive mood		imperative
SEQ	Sequential aspect	TOP	topicaliser

## 11.2 Orthography and pronunciation

Forms in this study are given using the practical orthographies adopted for northern Vanuatu languages. Many conventions are unproblematic, and reflect their expected phonetic value: this is the case of *p*, *t*, *k*, *l*, *r*, *m*, *n*, *s*, *h*, *w*, as well as *a*, *i*, *u*, etc.

Several conventions are shared throughout the region:

- *g* is a voiced velar fricative [χ], realised as approximant [w] syllable-finally
- *ñ* is [ŋ]; *ṁ* is [ŋmʷ]; *j* is [f]; *y* is [j]; *v* is [β]~[v]

Some conventions are specific to some languages:

- *b*, *d*, *ḡ* represent prenasalised stops [ʷb], [ʷd], [ʷg]; but *d* is a voiceless laminal stop [t̪] in Lo-Toga
- *q* is [kʷ] in Hiw, Lo-Toga, Lehali and Mwerlap; [kpʷ] elsewhere
- *q̄* in Volow is [ŋgbʷ]; *r̄* in Hiw is [ḡL]
- in the Banks, *e* is [ɛ]; *ē* is [œ]
- in the Torres, *e* is [ə]; *ē* is [ɛ] in Lo-Toga, [e] in Hiw
- *ē* is [e] in Lo-Toga, [i] everywhere else
- *ō* [ɔ] contrasts everywhere with *o* [ɔ]
- *ö* is [n] in Lehali, [ø] elsewhere
- *ā* is [n] in Lemerig, [a:] in Dorig
- *ā* is [ɛā] in Koro, [æ] elsewhere

These rules can be illustrated with some of the directionals given in Table 6 p.16. Thus *ma* is [ma]; *me* is [mə] in Hiw and Lo-Toga, but [mɛ] elsewhere; *mē* is [mɪ]. Lo-Toga *vēn* is [ben]; in Hiw, *vēn* is [βm̩] ‘up’ and contrasts with *vēn* [ben] ‘thither’. Likewise, *how* is [həw], *hōw* [həw], *sōw* [səw], *suwō* [suwɔ]; *sag* [saŋ], *seag* [sɛaŋ], *hag* [haŋ], *ag* [aŋ]; *vēn* [βm̩]; *wēl* [wəl], *wōl* [wɔl]; *row* [rəw], *rōw* [r̄ɔw]; *pāh* [pəh].

## 11.3 Etymological notes

Even though this study intends to describe the paradigms of space directionals following a synchronical approach, reference is occasionally made to etymologies. This appendix recapitulates what is known of the origin of modern directionals, based on the knowledge of regular correspondences in the area (François 2005, f/c). All the forms mentioned here were presented in Table 6 p.16; their phonetic transcription was given in §11.2.

### 11.3.1 Hither

All northern Vanuatu languages encode ‘hither’ using a regular reflex of POC \*mai:

- (I) POC \*mai ‘come; hither’: HIW *me*; LTG *me*; LHI *ma*; LYP *me*; VLW *me*; MTP *me*; LMG *me*; VRA *ma*; VRS *me*; MSN *me*; MTA *ma*; NUM *ma*; DRG *ma*; KRO *ma*; OLR *ma*; LKN *ma*; MRL *mē*.

### 11.3.2 Thither

What I gloss ‘thither’ for a shortcut is the allocentric participant-oriented directional (§2.4.1); a longer gloss would be ‘towards non-speaker’, i.e. ‘towards you ~ him ~ her ~ it ~ them’.

Several modern forms reflect POc \**watu* '(go) towards addressee' (Ross 2003:279):

- (II) POc \**watu* 'go towards addressee' → PTB \**atu* 'towards non-speaker, *thither*':

VRS *n|et*; MSN *n|at*; MTA *at*; NUM *at*; DRG *āt*; KRO *āt*; OLR *at*; LKN *at*; MRL *ot*.

Other northern languages reflect POc \**pano* 'go away (from speaker)' (Ross 2007:279):

- (III) POc \**pano* 'go away (from speaker)' → PTB \**vano* 'towards non-speaker, *thither*':

HIW *vēn*; LTG *vēn*; LHI *van*; LYP *van*; VLW *va*; MTP *van*.

See the discussion in §8.2.4.

For the same meaning 'thither, towards non-speaker', Lemerig uses its directional *wēl* (also 'across'), and Vera'a its directional *suwō* (also 'down').

### **11.3.3 Across**

As discussed in §4.2.2 and §8.2, the same POc verb \**pano* is not only the source of the allocentric participant-oriented directional ('thither') in some languages, but also of the directional 'across' used on the transverse axis in the local scale (§4.2.2). While this meaning can be reconstructed as far back as Proto Oceanic (François 2004), it is only reflected in four languages of the Torres-Banks area:

- (IV) POc \**pano* 'move in transverse direction'

→ 'following a direction parallel to the shoreline':

HIW *vēn*; MTA *vano*; NUM *van*; MRL *van*.

The languages of Gaua have non-cognate directionals *vak* and *pāh*, of unknown origin.

Finally, the languages of Vanua Lava show evidence of a lexical replacement of \**pano* with a local etymon \**volo* 'crosswise, across' (François f/c:195):

- (V) PTB \**volo* 'crosswise, across' → 'following a direction parallel to the shoreline':

LMG *wēl*; VRA *wōl*; VRS *wōl*; MSN *wol*.

The latter etymon later underwent semantic narrowing to 'parallel to shoreline towards southeast (for short distances on land)' – see the discussion in §5.2.2.

### **11.3.4 Up**

The following forms reflect Proto Oceanic \**sake* 'go up; up' (and related meanings):

- (VI) POc \**sake* 'go up; up': HIW *ag*; LTG *ag* ~ *iag*; LHI *ha*; LYP *sa*; VLW *ha*; MTP *hag*;

LMG *sag*; VRA *sag*; VRS *siag*; MSN *sag*; MTA *sage*; NUM *sa*; DRG *sag*;

KRO *sa* ~ *sag*; OLR *saa*; LKN *hag, rok|a*; MRL *seag*.

Three languages have created a new directional for 'up', from a verb 'climb' which can be reconstructed as \**vene* at the level of Proto Torres-Banks (PTB):

- (VII) PTB \**vene* 'climb' → 'up': HIW *vēn*; LTG *vin*; LHI *vēn*.

This process of lexical replacement, and its impact, are explained in §7 and 8.3.

### **11.3.5 Down**

The counterpart of \**sake* was \**sipo* in POc. However, only two northern Vanuatu languages show regular reflexes of \**sipo* in their directional form for 'down', namely Lo-Toga and Mota.

- (VIII) POc \**sipo* 'go down; down': LTG *iw*; MTA *swo* ~ *siwo*.

Other Torres-Banks languages reflect \**sipo* as a verb, but not as a directional. As for their directionals ‘down’, they point to a protoform \**suwo* or \**suwe*, of unknown origin. It may be an irregular reflex of POc \**sipo*, or reflect another lexeme.

- (IX) PTB \**suw[o,e]* ‘down’: HIW *uw*; LHI *how*; LYP *sōw*; VLW *hō*; MTP *hōw*; LMG *sōw*; VRA *suwō*; VRS *sōw*; MSN *sōw*; LKN *hōw*, *rōkōw*; MRL *sōw*.

Four Gaua languages have innovated a new directional ‘down’ from an etymon \**roro*, demonstrably a stative verb meaning originally ‘go deep, sink; be deep, be low’ (François 2010:139):

- (X) PTB \**roro* ‘sink, be deep, be low’ → ‘down’: NUM *ror*; DRG *ror*; KRO *ror*; OLR *roy*.

This case of total relexification is mentioned in §7.4.

### 11.3.6 In

Following the discussion in §2.4.2, it seems that Proto Oceanic did not have any lexemes for ‘in’ and ‘out’. These two directionals are thus local innovations only found in northern Vanuatu; they are absent from the conservative languages of Gaua.

Most Torres-Banks languages with a directional for ‘in’ reflect a local protoform \**saro*.<sup>29</sup>

- (XI) PTB \**saro* ‘enter, go in’ → ‘in’: LYP *say*; VLW *ha*; MTP *hay*; LMG *sar*; VRA *sar*; VRS *sar*; MSN *sar*; NUM *sar*; MRL *sar*.

This form \**saro* must have been originally a verb meaning ‘enter, go in’. Its suffixed form \**sarovayi* (reflecting the POc applicative \*-*akin*) is reflected in Mota and Mwotlap as a verb with the same meaning [cf. (12) for an example]:

- (XI') PTB \**sarovayi* ‘enter, go in’: MTP *hayveg*; MTA *sarovag*.

Three northern languages have innovated a different directional for ‘in’; these reflect an etymon \**ila*, of unknown origin:

- (XII) PTB \*\**ila* ‘??’ → ‘in’: HIW *iy*; LTG *il*; LHI *ila* ~ *la*.

### 11.3.7 Out

All northern Vanuatu languages unanimously reflect a protoform \**rowo* for ‘out’. This directional is most probably cognate with the verb \**rowo* in the same languages, meaning ‘dash, move swiftly, escape’, itself a regular reflex of POc \**Ropok* ‘dash, fly’.

- (XIII) POc \**Ropok* ‘dash, fly’ → PTB \**rowo* ‘dash, move swiftly, escape’ → ‘out’:

HIW *rōw*; LTG *rōw*; LHI *yow*; LYP *yow*; VLW *yo*; MTP *yow*; LMG *row*; VRA *rōw*; VRS *rōw*; MSN *row*; MTA *rowo*; MRL *row*.

This lexical innovation is not reflected in Gaua languages, which preserve the colexification of ‘out’ with ‘down’ inherited from Proto Oceanic (§2.4.2).

<sup>29</sup> In several languages (Löyöp, Vurës, Mwesen, Mwerlap), this directional ‘in’ is homophonous with the noun meaning ‘village clearing, dancing area in the centre of the village’. This similarity is purely accidental: the directional reflects \**saro*, whereas the noun reflects an etymon \**zara* ‘sweep, broom’ (Clark 2009:238).

## 12 References

- Ballu, Valérie, Marie-Noëlle Bouin, Patricia Siméoni, Wayne C. Crawford, Stéphane Calmant, Jean-Michel Boré, Tony Kanas, & Bernard Pelletier. 2011. Comparing the role of absolute sea-level rise and vertical tectonic motions in coastal flooding, Torres Islands (Vanuatu). *Proc. Nat. Acad. Sci. (PNAS)* 108 (32):13019-13022.
- Bennardo, Giovanni, ed. 2002. *Representing space in Oceania: Culture in language in mind*. Pacific Linguistics, 523. Canberra: Australian National University.
- Brown, Penelope, & Stephen Levinson. 1992. 'Left' and 'right' in Tenejapa: investigating a linguistic and conceptual gap. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 45:590-611.
- . 1993. 'Uphill' and 'downhill' in Tzeltal. *Journal of Linguistic Anthropology* 3:46-74.
- Cablitz, Gabriele. 2006. *Marquesan: A grammar of space*. New York: Mouton de Gruyter.
- Clark, Ross. 2009. *\*Leo Tuai: A comparative lexical study of North and Central Vanuatu languages*. Pacific Linguistics, 603. Canberra: Australian National University.
- Corbett, Greville. 2007. Canonical typology, suppletion and possible words. *Language* 83 (1):8-42.
- Danziger, Eve. 2011. Distinguishing three-dimensional forms from their mirror-images: Whorfian results from users of intrinsic frames of linguistic reference. *Language Sciences*.
- Dasen, Pierre, & Ramesh Mishra. 2010. *Development of Geocentric Spatial Language and Cognition: An Eco-Cultural Perspective*. Cambridge Studies in Cognitive and Perceptual Development, 12. Cambridge: Cambridge University Press.
- Dixon, R.M.W. 1988. *A Grammar of Boumaa Fijian*. Chicago: University of Chicago Press.
- François, Alexandre. 2003. Of men, hills and winds: Space directionals in Mwotlap. *Oceanic Linguistics* 42 (2):407-437.
- . 2004. Reconstructing the geocentric system of Proto Oceanic. *Oceanic Linguistics* 43-1:1-32.
- . 2005. Unraveling the History of the Vowels of Seventeen Northern Vanuatu Languages. *Oceanic Linguistics* 44 (2):443-504.
- . 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In *From Polysemy to Semantic change: Towards a Typology of Lexical Semantic Associations*, edited by M. Vanhove. Studies in Language Companion Series, 106. New York, Amsterdam: Benjamins. Pp.163-215.
- . 2009. The languages of Vanikoro: Three lexicons and one grammar. In *Discovering history through language: Papers in honour of Malcolm Ross*, edited by B. Evans. Pacific Linguistics, 605. Canberra: Australian National University. Pp.103-126.
- . 2010. Des valeurs en héritage: Les isomorphismes sémantiques et la reconstruction des langues. In Injoo Choi-Jonin, Marc Duval & Olivier Soutet (eds), *Typologie et comparatisme. Hommages offerts à Alain Lemaréchal*. Orbis-Supplementa 29. Louvain: Peeters. Pp.129-145.
- . 2011. Social ecology and language history in the northern Vanuatu linkage: A tale of divergence and convergence. *Journal of Historical Linguistics* 1 (2): 175-246.
- . 2012. The dynamics of linguistic diversity: Egalitarian multilingualism and power imbalance among northern Vanuatu languages. *International Journal of the Sociology of Language* 214: 85-110.
- . forthcoming. Shadows of bygone lives: The histories of spiritual words in northern Vanuatu. In *Lexical and structural etymology: Beyond word histories*, edited by R. Mailhammer. Studies in Language Change. Berlin: DeGruyter Mouton.
- Haviland, John. 1993. Anchoring, iconicity and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology* 3 (1):3-45.
- Hyslop, Catriona. 2002. Hiding behind trees on Ambae: Spatial reference in an Oceanic language of Vanuatu. In Bennardo (ed.), 47-76.
- Kalyan, Siva & François, Alexandre. f/c. Freeing the Comparative Method from the tree model: A framework for Historical Glottometry. In *The tree model*, edited by R. Kikusawa. [publisher TBA]
- Levinson, Stephen. 1992. Language and cognition: the cognitive consequences of spatial description in Guugu Yimithirr. In *Working paper no. 13*: Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics.
- . 1996a. Frames of reference and Molyneux's question: Crosslinguistic evidence. In *Language and space*, edited by P. Bloom, M. Peterson, L. Nadel & M. Garrett. Cambridge, MA: The MIT Press. Pp.109-170.
- . 1996b. Language and space. *Annual Review of Anthropology* 25:353-382.

- . 2003. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Language, Culture and Cognition, 5. Cambridge: Cambridge University Press
- Levinson, Stephen, & Penelope Brown. 1993. 'Uphill' and 'downhill' in Tzeltal. *Journal of Linguistic Anthropology* 3:46-74.
- Lichtenberk, Frantisek. 1983. *A grammar of Manam*. Oceanic Linguistics Special Publications, 18. Honolulu: University of Hawaii Press.
- Ozanne-Rivierre, Françoise. 1999. Spatial Orientation in some Austronesian Languages. In *Language Diversity and Cognitive Representations*, edited by C. Fuchs & S. Robert. Amsterdam/Philadelphia: John Benjamins. Pp.73-84.
- Palmer, Bill. 2002. Absolute spatial reference and the grammaticalisation of perceptually salient phenomena. In Bennardo (ed.), 107-157.
- . 2007. Pointing at the lagoon: Directional terms in Oceanic atoll-based languages. In *Language Description, History and Development: Linguistic Indulgence in Memory of Terry Crowley*, edited by J. Lynch, J. Siegel & D. Eades. Creole Language Library, 30. New York: John Benjamins. Pp.101-118.
- Paviour-Smith, Martin. 2009. Up and down in Aulua. *Te Reo* 52:47-72.
- Pederson, Eric, Eve Danziger, David Wilkins, Stephen Levinson, Sotaro Kita, & Gunter Senft. 1998. Typology and Spatial Conceptualization. *Language* 74 (3):557-589.
- Ross, Malcolm. 1988. *Proto-Oceanic and the Austronesian languages of Western Melanesia*. Pacific Linguistics. Canberra: Australian National University.
- . 2004. Demonstratives, local nouns and directionals in Oceanic languages: a diachronic perspective. In *Deixis and demonstratives in Oceanic languages*, edited by G. Senft. Pacific Linguistics, 562. Canberra: Australian National University. Pp.175-204.
- . 2007. Talking about space: terms of location and direction. In *The lexicon of Proto Oceanic: The physical environment*, vol. 2, edited by M. Ross, A. Pawley & M. Osmond. (First edition: 2003). Pacific Linguistics 545. Canberra: Australian National University. Pp.229-294.
- Senft, Gunter (ed.). 1997. *Referring to space. Studies in Austronesian and Papuan languages*. Oxford Studies in Anthropological Linguistics. Oxford: Oxford University Press.

# Freeing the Comparative Method from the tree model

*A framework for Historical Glottometry*

Siva Kalyan

*Northumbria University  
— Australian National University*

Alexandre François

*CNRS LACITO; Labex Empirical Foundations of Linguistics  
— Australian National University*

## Abstract

*Since the beginnings of historical linguistics, the family tree has been the most widely accepted model for representing historical relations between languages. While this sort of representation is easy to grasp, and allows for a simple, attractive account of the development of a language family, the assumptions made by the tree model are applicable in only a small number of cases: namely, when a speaker population undergoes successive splits, with subsequent loss of contact among subgroups. A tree structure is unsuited for dealing with dialect continua, as well as language families that develop out of dialect continua (for which Ross 1988 uses the term “linkage”); in these situations, the scopes of innovations (in other words, their isoglosses) are not nested, but rather they persistently intersect, so that any proposed tree representation is met with abundant counterexamples. In this paper, we define “Historical Glottometry”, a new method capable of identifying and representing genealogical subgroups even when they intersect. Finally, we apply this glottometric method to a specific linkage, consisting of 17 Oceanic languages spoken in northern Vanuatu.*

## 1. Introduction

The use of genealogical trees for the representation of language families is nearly as old as the discipline of historical linguistics itself; it was first proposed by August Schleicher in 1853, six years before Darwin proposed a tree model in evolutionary biology (e.g. Minaka & Sugiyama 2012: 177). It has since been the dominant method of visualis-

ing historical relationships among languages, and for good reason: its simple structure allows any hypothetical representation of a language family to be interpreted unambiguously as a set of claims about the sequence of demographic and social events that actually occurred in the history of the communities involved. These hypotheses can then potentially be falsified by new data or analysis, leading to a more valid representation. Other methods of representing the historical relationships among languages have from time to time been proposed and defended—e.g. Johannes Schmidt's (1872) “Wave Model”, Southworth's (1964) “tree-envelopes” (akin to the “population trees” used in phylogeography, e.g. Avise 2000: 32), Anttila's (1989: 305) isogloss map, Hock's (1991: 452) “truncated octopus'-like tree”, van Driem's (2001) “fallen leaves”, and most recently NeighborNet (Bryant *et al.* 2005), among many others. However, to our knowledge, none of these has combined precision and formalisation with direct interpretability in terms of historical events, to the extent that has been achieved by the family-tree model.<sup>1</sup>

Yet there are important reasons to be dissatisfied with the family-tree model (as has frequently been pointed out; see also Bloomfield 1933: §§18.9–12). In particular, it rests entirely on the assumption that the process of language diversification is one where language communities undergo successive splits—via migration or other forms of social disruption—with subsequent loss of contact. While this particular social scenario may have occurred occasionally (e.g. in the separation of Proto-Oceanic from the remainder of the Austronesian language family; see Pawley 1999), it can hardly be regarded as the general case.

The way language change arises is via a process of *language-internal diffusion* (François forthc.; cf. Labov 1963, Milroy & Milroy 1985, Croft 2000: 166–195; Enfield 2008)—as speakers in a network imitate each other so as to jointly adopt an innovative speech habit. When the innovation settles into a certain section of the social group, it becomes part of its linguistic heritage and can be transmitted to its descendants. This diffusion process is the underlying mechanism behind “genetic” relations (or better, to use Haspelmath's (2004:222) preferred term, “genealogical” relations) among languages, whereby each subgroup is defined by the innovations its members have undergone together. Whereas contact-induced change takes place between separate languages, the process of *language-internal diffusion* that defines language genealogy involves mutually intelligible speech varieties.

The tree model can represent genealogical relations in just one particular case: when a language community has split into separate groups, each of which later goes through its own innovations. But this model cannot properly handle the frequent case when adjacent speech communities remain in contact even after undergoing innovations that increase their difference. In such situations, provided the speech varieties remain mutually intelligible for some time, nothing prevents successive innovations from targeting overlapping portions of the network: e.g. one isogloss targeting dialects A-B-C, another one C-D-E, then B-C, then D-E-F, etc. In such cases of dialect chains or networks, frequently observed in dialectology (and described further below), the layering of partially overlapping innovations results in *intersecting* genealogical subgroups—a situation which cannot be addressed by the tree model (Gray *et al.* 2010:3229).

<sup>1</sup> The authors wish to thank Malcolm Ross, Mark Donohue and Martine Mazaudon for their comments on an earlier draft of this paper. They would also like to thank the other participants of the

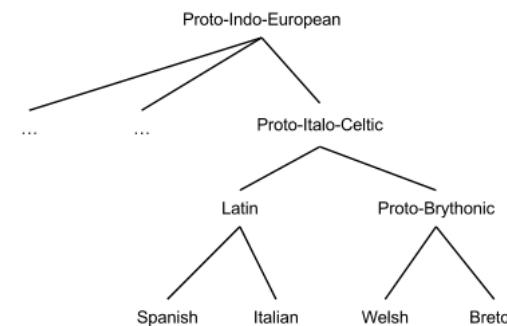
As is increasingly evident from the work of a number of historical linguists, this sort of intersecting configuration typical of dialect continua is also the normal situation in most language families around the world (e.g. Geraghty 1983; Ross 1988; Toulmin 2009; Heggarty *et al.* 2010; Huehnegaard & Rubin 2011). Of course, any set of data could be forced into a tree structure, but in most cases this can only be done by selectively discarding some of the data—no doubt in good faith—so as to retain only those which are compatible with a particular subgrouping hypothesis. Debates about which tree best represents the language family thus usually boil down to (often pointless) arguments over which parts of the data may be ignored.

In this study, we start by elaborating on the arguments and the claims made in the preceding paragraphs, by illustrating in greater detail how trees are used in historical linguistics, and discussing their advantages and disadvantages. We then move to the task of proposing a new method of representing genealogical relationships among languages, which we call Historical Glottometry. While ultimately inspired by the Wave Model which Schmidt (1872) proposed as an alternative to the family tree, our method also draws on the quantitative approach of dialectometry (Séguy 1973; Goebel 2006; Szemrešányi 2011). We hope this model provides more realistic insights into language history than the tree model, while still combining precision and formalisation with historical interpretability. Finally, we illustrate our model by applying it to a group of seventeen Oceanic languages spoken in Vanuatu, an archipelago in the south Pacific.

## 2. Subgrouping in the tree model

### 2.1. An example from Indo-European

Consider the family tree shown in Figure 1, which represents a selection from the family of Indo-European languages. At the bottom are languages that are currently spoken; languages higher in the tree are ancestors of the languages that branch from them. Each nodal ancestor is called a proto-language, whose descendants together form a subgroup.

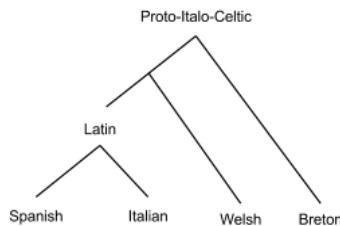


**Figure 1:** A selection of Indo-European languages, organised as a tree.

In some cases, ancestor languages have been preserved in writing; thus we have direct evidence that (some variety of) Latin is the common ancestor of Spanish and Italian. In other cases, the ancestors are hypothetical, and must be reconstructed by comparing their surviving descendants; thus it is merely a hypothesis that there was a unified Proto-Brythonic language from which Welsh and Breton are descended, and the features of this proto-language are also hypothetical.

Ancestral languages (whether attested or reconstructed) can themselves be compared, and their own ancestors hypothesised and reconstructed, in a recursive fashion. Thus, some linguists (e.g. Kortlandt 2007) believe that Latin and Proto-Brythonic ultimately descend from a language termed Proto-Italo-Celtic (PIC).<sup>2</sup> Repeatedly applying this process of comparison and reconstruction—called the *Comparative Method*—leads to proto-languages further and further back in time, ultimately ending in Proto-Indo-European (PIE).<sup>3</sup>

Granted that the uppermost node, Proto-Indo-European, is valid (since the Indo-European languages are indeed related to one another), on what basis are lower-level proto-languages (or equivalently, subgroups) posited? For example, why isn't Welsh grouped with Latin, separately from Breton, as in the fictitious Figure 2?



**Figure 2:** An incorrect tree of Italo-Celtic languages.

The reason is that this would imply that Latin and Welsh both exhibit certain changes (or *innovations*) from PIC (and hence, from PIE) that are not exhibited by Breton. But there are no notable innovations of this kind. Also, Figure 2 would imply that there are *no* innovations shared by Welsh and Breton which are not also shared by Latin (and all other members of the Italo-Celtic subgroup). This too is false: for example, the Brythonic languages changed \**k*ʷ to *p*, and changed \**s* to *h* at the beginnings of words (Schmidt 2002: 80–81); Latin, on the other hand, preserved these sounds intact. In sum, the representation in Figure 1 is more faithful to the empirical data we have from attested languages, than is Figure 2.

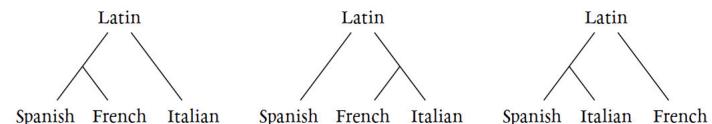
As we have just illustrated, in the Comparative Method, a subgroup is posited on the basis of *exclusively shared innovations* among its members—a principle first formulat-

<sup>2</sup> Brythonic is actually a branch of Celtic, which in turn is a branch of Italo-Celtic; likewise, Latin is a member of the Italic branch of Italo-Celtic. The fact that the existence of Proto Italo-Celtic is controversial is irrelevant to the present demonstration—what is important is that Latin and the Brythonic languages do in fact have a common ancestor (even if that ancestor turns out to be nothing other than Proto-Indo-European itself).

<sup>3</sup> On general principles of the comparative method, see Hock (1991), Campbell (2004), Crowley & Bowern (2010), among many others.

ed by Leskien (1876: xiii). In other words, a subgroup represents a hypothesis that all of its members share certain innovations that are not exhibited by any other language, and that any innovation that a member shares with a non-member is necessarily shared by *all* members. (This is similar to how, in phylogenetics, clades are interpreted as monophyletic groups defined by synapomorphies: see Skelton et al. 2002: 27–28.)

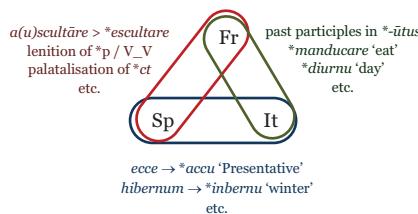
Let us now consider what happens when we add another language—French—to our tree. There is no question but that French is a descendant of Latin; hence it should ultimately be a daughter of the “Latin” node. However, there are multiple ways in which it could be put into a tree together with Spanish and Italian (Figure 3). Which of these choices is correct?



**Figure 3:** Three possible ways to represent the relations between Spanish, French and Italian.

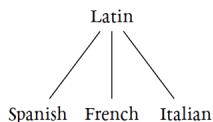
Choice 1, with Spanish and French forming a subgroup, seems justified by the innovations that are shared between these two languages, and not shared by Italian: for example, the irregular change of *a(u)scultāre* ‘listen’ to *\*escultāre* > Sp. *escuchar*, Fr. *écoutier*, vs. It. *ascoltare* (Berger & Brasseur 2004: 90); intervocalic lenition of *\*p*—e.g. *ripa* ‘riverbank’ > Sp. *riba*, Fr. *rive*, vs. It. *ripa* (Posner 1996: 234); and the palatalisation of *\*ct* clusters—e.g. *factum* ‘done’ > Sp. *hecho*, Fr. *fait* vs. It. *fatto* (Hall 1950: 25). However, one can also find innovations shared by French and Italian but not by Spanish, which would argue in favour of choice 2: for example, the innovative weak past participle suffix *\*-ūtus* which affected many verbs—e.g. *\*sapūtus* ‘known’ > It. *saputo*, Fr. *su*, as opposed to Sp. *sabido* < *\*sapūtus* (Alkire & Rosen 2010: 177); or numerous lexical innovations such as *\*diurnu* > It. *giorno*, Fr. *jour* ‘day’, replacing Lat. *dīes* (Sp. *día*), or *\*manducāre* ‘chew’ > It. *mangiare*, Fr. *manger* ‘eat’, replacing Lat. *comedere* (Sp. *comer*). Finally, one could cite evidence in favour of subgrouping Spanish and Italian together as opposed to French (as in choice 3), e.g. the irregular change of Lat. *ecce* to *\*accu* (Wüest 1994), as in the (feminine) distal demonstrative *\*accu-illa* > Sp. *aquella*, It. *quella*, where French preserves *ecce* (*\*ecce-illa* > Fr. *celle*); or the irregular insertion of /n/ in *hibernum* ‘winter’, yielding *\*inbernu* > Sp. *invierno*, It. *inverno*, vs. Fr. *hiver* (Alkire & Rosen 2010: 339). Many other examples of exclusively shared innovations<sup>4</sup> could be found for each of the three language pairs. In all cases, the nature of the changes (especially phonological and morphological change, whether regular or irregular) is typical of the sort of evidence that is traditionally considered diagnostic of genealogical subgroups under the Comparative Method.

<sup>4</sup> Obviously, the term “exclusively” must be understood within the restricted set of three languages taken here for the sake of discussion. Some of the innovations shared by French and Spanish are also shared with Catalan, Portuguese, etc., but this is not relevant for the present demonstration. (Interestingly, Catalan seems to exhibit most of the innovations mentioned.)



**Figure 4:** Historical evidence supports three intersecting subgroups involving Spanish, French and Italian—a situation incompatible with the family tree model.

In this particular case, the data simultaneously support three intersecting subgroups (Figure 4): Spanish–French, French–Italian and Spanish–Italian. The tree model, which would force us to privilege one of these three groupings at the expense of the other two,<sup>5</sup> is unable to do justice to the empirical evidence.



**Figure 5:** A rake (or “polytomy”).

One could be tempted to represent this thorny situation by resorting to the diagram in Figure 5, which does not necessarily commit us to any subgrouping hypothesis. This sort of diagram (cf. Ross 1997: 213) is sometimes used as an “agnostic” representation, which Pawley (1999) calls a “rake-like” structure, and van Driem (2001) likens to “fallen leaves”. (In phylogenetics this is known as “(soft) polytomy”;<sup>6</sup> see Page & Holmes 2009: 13.) Yet it too is unsatisfactory, as it could be interpreted as claiming that there are *no* exclusively shared innovations between Spanish and French, between French and Italian, or between Spanish and Italian, when—as we have seen—there is in fact solid, positive evidence for all of these. (In phylogenetic terms, a rake is ambiguous between “soft polytomy” and “hard polytomy”.) Even if we specifically exclude this latter interpretation, we are only left with the impression that science is simply incapable of unraveling the precise linguistic history of the language family. While this is sometimes the case due to lack of data, it is certainly not the case in such a well-documented family as Romance.

The history of individual changes across Romance dialects and languages is extremely well-known: if this family cannot be represented by a tree, then this cannot be

<sup>5</sup> This is what Hall (1950) does: his assumption that languages must evolve following a cladistic model has him force the data into a tree structure. His “Western Romance” node, by grouping French and Spanish together, arbitrarily favours only one of the three groupings outlined here, and deliberately ignores any conflicting evidence.

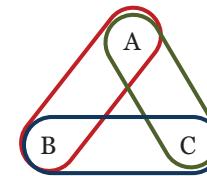
<sup>6</sup> We are grateful to Nobuhiro Minaka (p.c.) for pointing this out.

due to a lack of data, but to the inherent flaws of the tree model itself: in particular, the axiom that genealogical subgroups defined by exclusively shared innovations are necessarily nested, and never intersect. This axiom results from an incorrect understanding of language change (cf. Bossong 2009, François forthc.), namely that an innovation consistently results in total social isolation and lack of contact with communities that did not undergo the innovation—an incorrect assumption in most of the world’s history. What we see, on the contrary, is that the spread of an innovation within part of a dialect network, insofar as it still allows mutual intelligibility with non-participating dialects, can perfectly well be followed by other innovations whose geographical scope may cross-cut its own, resulting in intersecting subgroups. We need a model of language relationships that is capable of accommodating such situations in a more accurate and faithful way than the tree model.

## 2.2. The problem of linkages

We can generalise our observations above by considering an abstract case, consisting of a family of three languages: A, B, and C. If A and B have some exclusively shared innovations, but neither B and C nor A and C do, then the situation is amenable to a tree representation (as in choice 1 in Figure 3 above). Historically, this represents a situation where the Proto-ABC speech community somehow split into two groups, one of which (the common ancestor of the modern A and B communities) underwent certain linguistic innovations, separately from C; these innovations are said to have resulted in a hypothetical language “Proto-AB”. Later on, a similar split took place in the Proto-AB community, that resulted in the separate development of A and B.

But another situation is also possible, as we saw in the case of Romance languages. This is the case where there are exclusively shared innovations not only between A and B, but also between B and C, and/or between A and C: that is, a situation in which shared innovations define intersecting groupings—see Figure 6 (and Figure 4 above).



**Figure 6:** When shared innovations intersect

This situation cannot be represented using the tree model, which assumes that a language can belong to one genealogical subgroup only. The only way to force the data into a tree—and posit, for example, a subgroup AB—would be to disregard the other two sets of innovations which contradict this grouping. Admittedly, such a procedure may be tenable in some cases. For example, C could have undergone some of the same innovations as A and B purely by chance, so that these are not really “shared innovations” in the relevant sense, but are rather “parallel innovations”. The trouble with this argument is that it is often extremely difficult to come up with positive evidence for it. In particular, if it is believed that C was still in contact with A and B at the time it underwent these innovations, it is unparsimonious to invoke independent, parallel devel-

opment as an explanation: it is more probable that the changes they have in common reflect events of language-internal diffusion across dialects.

Another situation in which it may be reasonable to disregard the B–C and A–C innovations is when there is good reason to believe that these all occurred historically *after* the A–B innovations, and at a point in time when C had already become mutually unintelligible with A and B (i.e. had become a separate language). In this case, many historical linguists would label the B–C and A–C innovations as effects of “language contact” (or “horizontal transmission”), and would disregard them for the purpose of representing genealogical relationships. This sort of reasoning only works under the assumption that it is possible to draw a principled line between diffusion across language boundaries (“contact”) and diffusion within them (“internal change”). This seems unlikely, given that the concept of a “language boundary” (i.e. whether two speech varieties are separate languages or simply dialects of the same language) is itself a gradient notion. However, the argument of contact is usually proposed in good faith, and may be accepted in some obvious cases, namely when the genealogical distance between the speech varieties involved was already much too great at the time of contact for mutual intelligibility—e.g. lexical borrowings from Old Norse into Old English, or from Polynesian languages into other Oceanic languages (Biggs 1965).

In sum, given a set of changes with overlapping distributions, there are occasionally *bona fide* reasons for arguing that some of them are *not* genealogical in nature, and thus should be discarded for the purpose of subgrouping. In general, though, there is often no legitimate basis for deciding which ones may be ignored. Sometimes, this is merely due to lack of evidence (historical or linguistic) about which set of changes predates the other. But in many cases, the problem is simply that the tree model fails to capture the fact that innovations do spread in entangled patterns across sets of mutually intelligible dialects, resulting in intersecting genealogical subgroups. This is what happens in dialect chains and networks, as well as in full-fledged language families that have evolved out of dialect networks—which Ross (1988:8; 1997:213) calls *linkages*. The relationships among Spanish, French and Italian—or among other Romance languages, for that matter (with the possible exception of Romanian)—are typical of a linkage. Crucially, linkages are common throughout the world: similar configurations have been described—under various names—for Sinitic (Hashimoto 1992; Chappell 2001), Semitic (Huehnergard & Rubin 2011), Indo-Aryan (Toulmin 2009), Athabaskan (Krauss & Golla 1981; Holton 2011), Oceanic (Geraghty 1983, Ross 1988), and many other language families. In Section 4, we will be presenting a detailed example from a section of the Oceanic linkage.

In the case of linkages, decisions about which innovation-defined groupings should be ignored for the purpose of representing genealogical relationships tend to be *ad hoc*, and debates rage with no sign of resolution. In our view, such problems are mere artefacts of the assumptions present in the tree model, and lack any legitimate basis as far as language change is concerned. In fact there is no justification to the assumption that dialects and languages evolve primarily by splitting in a tree-like fashion: the more is known about language change, the more it becomes obvious that this model is a poor approximation of reality, and rests on a misleading metaphor.

In the remainder of this paper, we advance a more flexible model: Historical Glottometry. It elaborates on the principles of the Comparative Method, yet attempts to liberate it from the misleading influence of the family-tree model, by proposing a representation that reflects historical reality more faithfully.

### 3. Defining Historical Glottometry

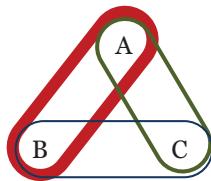
#### 3.1. Intersecting subgroups

Insofar as our model is meant to describe (past or present) dialect networks, it is useful to start by looking at how these are represented by dialectologists. A key concept in dialectology is that of the *isogloss* (Chambers & Trudgill 1998: 89). Considering a given linguistic property, and the way it is geographically distributed across a dialect network, an isogloss is a line delimiting the set of dialects (or of “communalects”, to use a term neutral between “language” and “dialect”) that share that property. Isoglosses can be represented on geographically realistic maps, or on more abstract figures. The lines in Figure 4 above are examples of isoglosses, showing the distribution of certain linguistic properties in (part of) the Romance family.

In principle, an isogloss may involve any property that is shared among languages, regardless of its historical origin. And indeed, because dialectology traditionally examines modern speech varieties from a purely synchronic perspective, isogloss maps often fail to distinguish between those similarities that result from shared innovations (*synapomorphies*) and those that are simply shared retentions from a common ancestor (*symplesiomorphies*), or even parallel innovations (*homoplasies*) and accidental similarities.<sup>7</sup> From the perspective of historical linguistics that concerns us here, it is indispensable to restrict our observations to *shared innovations*: indeed, as per Leskien’s principle mentioned above, it is a pillar of the Comparative Method that only innovations are indicative of the shared history of communities. The methodology we propose here can be seen as exactly this: a dialectological approach to language history, combining the precise descriptive tools of dialectology and dialectometry (Goebel 2006, Nerbonne 2010, Szemrecsányi 2011) with the powerful concepts of the Comparative Method—notably the stress on shared innovations.

One problem with isogloss maps (and admittedly the main reason why they have not been adopted more widely outside of dialectology) is that they become visually messy very quickly as more and more intersecting isoglosses are added; furthermore, they do not lend themselves to straightforward storytelling as much as a tree diagram would. The former issue, at least, can be addressed if we choose to use isoglosses to represent not individual innovations, but rather language groupings defined by one or more exclusively shared innovations (in other words, subgroups, in our extended sense of the term). A subgroup is simply a grouping of dialects or languages identified by a bundle of (innovation-defined) isoglosses. The thickness of the isogloss line can then be used to represent the strength of the evidence for each language grouping. For example, Figure 7 translates visually the fact that, while the three subgroups AB, AC and BC are all empirically supported, BC is the weakest pairing, and AB the strongest.

<sup>7</sup> Important exceptions include the “dialect map of the Indo-European languages” in Anttila (1989: 305), which is extremely similar in spirit to the model we will be proposing below, as well as the diagrams in Southworth (1964), which are less so. We are grateful to Malcolm Ross for having brought these works to our attention.



**Figure 7:** A representation of intersecting subgroups with relative weighting

With such a configuration of the data, historical linguists who take the tree model for granted might be tempted to favour AB as the only valid subgroup, and dismiss the evidence for the two other subgroups altogether, under the assumption that these “weaker” groupings must be mere illusions—whether their similarities be due to “contact”, or to “parallel innovation”, etc. However, unless there is indeed a principled way of ruling out these isoglosses, it is wiser to keep them in the picture: the idea is that those innovations that are shared between A and C, or B and C, reflect historical events of shared linguistic development just as much as do those between A and B. It is just that the social relations between communities A and B, over the entire course of the history of the ABC family, have been stronger, more frequent or more sustained than those between other pairs of communities. Historical Glottometry can be used precisely as a means to explore and evaluate the strengths of historical connections between social groups, based on the linguistic traces they left in modern languages.

In sum, linguistic linkages make it necessary to accept the idea of a language family in which genealogical subgroups have different strengths, and can cross-cut. Rather than a simplistic binary answer (*X forms vs. does not form a subgroup with Y*), subgrouping studies should allow for the possibility of *stronger vs. weaker subgroups*. Just as a village A may have more frequent mutual interaction with another village B than with C, likewise languages A and B can be said to form a stronger subgroup together (i.e., be “more subgroupy”) than languages A and C. Ideally, such claims could even be quantified—as in “A subgroups  $n$  times as strongly with B as it does with C”<sup>8</sup>.

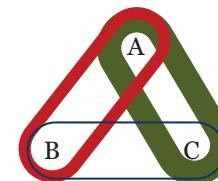
The crucial question is now: how can we define, and calculate, the “strength” of a subgroup? This is the object of the next subsection.

### 3.2. The cohesiveness of subgroups

The most obvious way to represent the strength of a subgroup using isoglosses would be to simply make the thickness of isoglosses directly proportional to the number of innovations defining the respective groupings. For example, suppose that in the above example of languages A, B and C, there were 12 innovations exclusively shared between A and B, 4 between A and C, and 2 between B and C: then our diagram would look exactly as in Figure 7 (where 1 shared innovation = 1 pixel).

<sup>8</sup> A further extension of our model, which we will not have room to develop in this study, could be to provide both *quantification* and *qualification* to genealogical relations. Thus one could imagine statements along the lines of “A subgroups with B twice as strongly as it does with C as far as regular sound change is concerned; but it does so 1.6 times more with C than with B with respect to verbal morphology, 3 times with respect to lexical replacement in basic vocabulary”, etc.

However, suppose that instead of 4 exclusively shared innovations between A and C, there were 24. Our diagram would then be as in Figure 8:



**Figure 8:** Intersecting isoglosses, with more support for AC than for AB.

Insofar as the thickness of their lines is exactly proportional to the number of exclusively shared innovations between each pair of languages, Figures 7 and 8 are accurate, fully-detailed representations of their respective data. However, they fail to represent an important fact: that the strength of the AB grouping in the first situation is *greater*, relative to the other isoglosses, than the strength of the same grouping in the second situation—despite the fact that the same number of defining innovations ( $n = 12$ ) is involved in both cases.

Interestingly, Pawley (2009: 13), discussing the factors that provide evidence for a particular subgrouping hypothesis, notes that “The weight of this evidence depends on the number and quality of the innovations concerned *and on the number and quality of innovations that have conflicting distributions*” (our emphasis). We thus need to quantify the strengths of groupings in a way that takes into account not only the absolute number of innovations that **support** the grouping, but also the number that **conflict** with it. An isogloss  $x$  is said to “conflict” with a subgroup  $y$  if they cross-cut each other—i.e. if and only if  $x$  contains some but not all members of  $y$ , and also contains members outside  $y$  (mathematically speaking:  $x \cap y$ ,  $x \setminus y$  and  $y \setminus x$  are all nonempty). In our case, even though the AB grouping is supported by 12 innovations in both cases, it is *more strongly* supported in the first case (where the 12 innovations of AB conflict with only 4 isoglosses for AC plus 2 for BC) than in the second (where the number of conflicting isoglosses is 24 + 2).

In the spirit of procedures common in Social Network Analysis (see Valente 1995, Carrington *et al.* 2005), we propose to define the “**cohesiveness**” of a subgroup as the proportion of **supporting** evidence with respect to the entire set of relevant evidence. Thus, for each given subgroup  $G$ , let  $p$  be the number of supporting innovations, and  $q$  the number of conflicting innovations. The total amount of evidence that is relevant for assessing the cohesiveness of  $G$  is  $(p + q)$ .<sup>9</sup> Now, if we call  $k_G$  the cohesiveness value of  $G$ , we have:

$$k_G = \frac{\text{number of supporting innovations}}{\text{total number of relevant innovations}} = \frac{p}{(p + q)}.$$

In the situation depicted in Figure 7, the cohesiveness of AB would be calculated as:

<sup>9</sup> Those innovations that are entirely nested within a subgroup (e.g. those that affected only the language B within AB, and no language outside AB) are irrelevant to the cohesiveness of that subgroup, and therefore do not take part in the calculations.

$$k_{AB} = \frac{12}{12+(4+2)} = \frac{12}{18} = \frac{2}{3} \approx 67\%.$$

This result can be translated into plain language by saying that, out of all the innovations that affected the subgroup AB (i.e. either encompassed the subgroup as a whole, or affected one of its members together with an external member), exactly two thirds confirmed the cohesion of AB as a subgroup, while one third contradicted it. More simply, A and B “moved together” two-thirds of the time, and “moved apart” one-third of the time.

In the situation depicted in Figure 8, the cohesiveness of AB would be:

$$k_{AB} = \frac{12}{12+(24+2)} = \frac{12}{38} \approx 32\%.$$

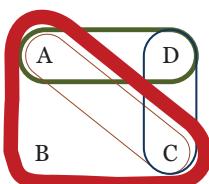
That is, in Figure 8, AB as a subgroup is confirmed 32% of the time, and contradicted 68% of the time.

These rates of 67% and 32% should be compared with the theoretical cohesiveness values which all subgroups are supposed to have in a “well-behaved” family tree, namely 100%. In an ideal tree, any group of languages defined by even a single shared innovation is supposed to *always* behave like a subgroup: that is, 100% of the innovations that affect it should confirm its cohesion, and there should be no genealogical innovation involving one (but not all) of its members together with a non-member. As we will see below with real data, this extreme figure of 100% is a convenient fiction that is virtually never met with among real-life languages—at least not in situations of linkages. Rates of cohesiveness in most subgroups typically fall far short of the “ideal” (in our data, most of them have a cohesiveness of between 10% and 30%). This does not mean that we are not dealing with genealogical subgroups at all; but rather, that this very notion must be redefined so as to accommodate the heterodox notion of the *strength* of a subgroup.

### 3.3. Subgroupiness

Given this measure of cohesiveness, we could use it to determine the thickness of our isogloss lines. However, cohesiveness alone is not sufficient to provide an accurate representation of each subgroup’s strength: as we will see now, it is necessary to also take into account the absolute number of exclusively shared innovations.

Consider now a family of four languages, A, B, C and D, where there are 12 innovations shared by ABC; 4 by AD; 2 by CD, and 1 by AC, as in Figure 9:



**Figure 9:** A family of four languages.

Note here that the number of innovations shared by AC ( $n = 1$ ) is irrelevant to the calculation of the cohesiveness of ABC, since it neither confirms this subgroup nor con-

tradicts it (see fn. 11). In order to assess the cohesiveness of ABC, what we need is to observe the number of innovations that confirm it ( $n = 12$ ) and those that clearly conflict with it—i.e. the innovations of AD ( $n = 4$ ) plus those of CD ( $n = 2$ ). The cohesiveness of ABC is thus:

$$k_{ABC} = \frac{12}{12+(4+2)} = \frac{12}{18} = \frac{2}{3} \approx 67\%.$$

Let us now calculate the cohesiveness of AC. This grouping is confirmed not only by the innovations that are exclusively shared by A and C ( $n = 1$ ), but also by those which they share non-exclusively, since these too show that languages A and C tend to undergo the same linguistic changes together. This includes, in Figure 9, the 12 innovations shared by ABC. As a result, the cohesiveness of the grouping AC should be like this:

$$k_{AC} = \frac{12+1}{(12+1)+(4+2)} = \frac{13}{19} \approx 68\%.$$

In sum, the cohesiveness of AC is even *greater* than that of ABC. Yet we would not want to say that AC is a “stronger” subgroup than ABC, because the latter has a far greater number of *exclusively* shared innovations.

Our proposed solution to this problem is to use the absolute number of *exclusively shared innovations* as the main point of reference, and qualify it using the subgroup’s cohesiveness rate ( $k$ ) as a **weighting** coefficient. For each given subgroup  $G$ , let  $\varepsilon$  be its number of exclusively shared innovations;  $p$  its number of supporting innovations (i.e. shared innovations, whether exclusively or not), and  $q$  the number of conflicting innovations. We already saw that the cohesiveness rate is  $k = \frac{p}{(p+q)}$ . We now propose to define the **subgroupiness** of a language cluster (call it ‘sigma’,  $\varsigma$ ) as the product of its cohesiveness rate ( $k$ ) with its number of exclusively shared innovations ( $\varepsilon$ ):

$$\varsigma = \varepsilon \times k = \varepsilon \times \frac{p}{(p+q)}.$$

For example, if we come back to the comparison of Figures 7 and 8, we can now weight the absolute number of innovations exclusively shared by A and B ( $\varepsilon_{AB}$ ) using AB’s cohesiveness rate  $k_{AB}$  (given above), and thus calculate its subgroupiness  $\varsigma_{AB}$ .

In Figure 7:

$$\varsigma_{AB} = 12 \times \frac{12}{18} = 8.$$

In Figure 8:

$$\varsigma_{AB} = 12 \times \frac{12}{38} \approx 3.79.$$

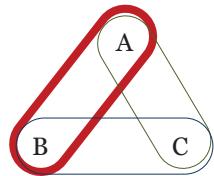
These numbers constitute exact measurements of the extent to which AB is a more strongly-supported subgroup in the first case than in the second case. (In other words, we can now say, “AB is more than twice as strongly supported—or more simply, *more than twice as subgroupy*—in Figure 7 than in Figure 8.”) As for Figure 9, we find that

$$\varsigma_{ABC} = 12 \times \frac{12}{18} = 8 \text{ and } \varsigma_{AC} = 1 \times \frac{13}{19} = \frac{13}{19} \approx 0.68,$$

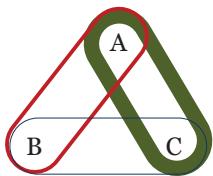
in other words, that ABC is more than eleven-and-a-half times as subgroupy as AC. These results are consistent with the intuition that the subgroup ABC is more strongly supported than AC. In conclusion, subgroupiness constitutes the best criterion we have found for assessing the relative strengths of the genealogical subgroups in a language family.

### 3.3. A visual representation

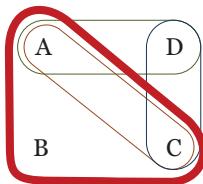
In terms of visual representation, it is then easy to draw lines around subgroups, whose thickness is proportional to their calculated subgroupiness  $\varsigma$ . Figures 7'–9' show our proposed representations of the situations depicted in Figures 7–9, respectively. We call these kinds of figures **historical glottometric diagrams** ('glottometric diagrams' for short).



**Figure 7'**: Illustration of subgroupiness-based isogloss thickness for the situation depicted in Figure 7. Subgroupiness rates:  $\varsigma_{AB} = 8$ ;  $\varsigma_{AC} = 0.89$ ;  $\varsigma_{BC} = 0.22$ .



**Figure 8'**: Illustration of subgroupiness-based isogloss thickness for the situation depicted in Figure 8. Subgroupiness rates:  $\varsigma_{AB} = 3.79$ ;  $\varsigma_{AC} = 15.16$ ;  $\varsigma_{BC} = 0.11$ .



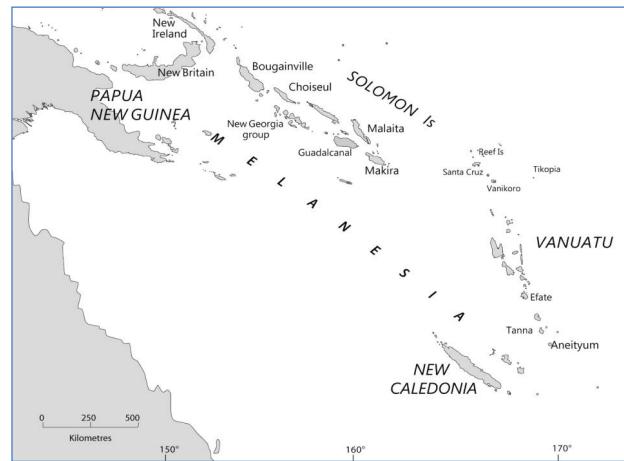
**Figure 9'**: Illustration of subgroupiness-based isogloss thickness for the situation depicted in Figure 9. Subgroupiness rates:  $\varsigma_{ABC} = 8$ ;  $\varsigma_{AD} = 0.84$ ;  $\varsigma_{AC} = 0.68$ ;  $\varsigma_{CD} = 0.21$ .

The examples given in this section were abstract, and simple in the sense that they involved small numbers of languages and of innovations. But the same tools can be profitably applied to a much richer set of data. The next section will show precisely how Historical Glottometry can be applied to a real dataset involving 17 languages, and a total of 474 innovations.

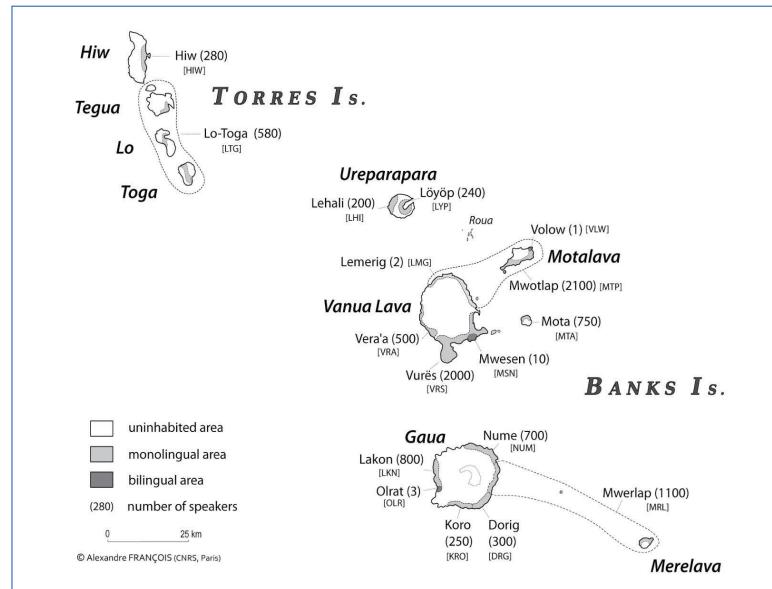
## 4. A case study from North Vanuatu

### 4.1. The languages

We can now illustrate the power of Historical Glottometry using a set of actual data from the languages of Vanuatu, an archipelago in the south Pacific (see Map 1).



**Map 1**: The archipelago of Vanuatu, in the South Pacific



**Map 2**: The 17 languages of the Torres and Banks Is., in northern Vanuatu

There are around 110 indigenous languages spoken in Vanuatu, which all belong to the Oceanic branch of the Austronesian language family. The evidence for Oceanic being a (classical, nearly 100% cohesive) subgroup of Austronesian is massive (Pawley &

Ross 1995; Ross 1988), and it is widely accepted that there was at some point a more-or-less homogeneous Proto-Oceanic language spoken throughout most of the south Pacific (Pawley & Green 1984; Pawley 2008, 2010), which gradually fragmented into dialects and then independent languages—following a scenario quite similar to the history of Romance languages. Over the decades, there have been a number of attempts to fit the modern-day languages of Vanuatu into a tree model. Clark (2009:4–9) lists as many as nine conflicting subgrouping hypotheses, none of which has reached consensus. This tends to confirm our hypothesis that the genealogical relations among Vanuatu languages cannot be rendered by a tree: they constitute a *linkage*, i.e. a group of modern languages which emerged through the *in situ* diversification of an earlier dialect network (Tryon 1996; François 2011a, 2011b).

We will be focusing on the two northernmost island groups of the Vanuatu archipelago, the Torres and Banks Islands. Alexandre François has been conducting fieldwork there since 1997, and has collected extensive data on the 17 languages still spoken in this small area, many of which are endangered (see François 2012). The names of these languages are given on Map 2, together with three-letter abbreviations and numbers of speakers.

#### 4.2. Intersecting isoglosses in North Vanuatu

The communalects of North Vanuatu have now lost mutual intelligibility, and constitute distinct languages. However, it is possible, thanks to the Comparative Method, to unravel the various linguistic changes that took place since the time of earlier linguistic unity, and brought about the present linguistic diversity (François 2005, 2011a, 2011b). Even though some changes affected a single communalect in isolation, the most typical case was for a given innovation to emerge in some location, and diffuse via social interaction from one dialect to its neighbours, until it settled down into a certain portion of the dialect network. Some isoglosses encompassed the entire area, while others only targeted a set of four or five villages. And of course, in a manner similar to Romance dialects, what we see is that the isoglosses defined by the various innovations cross-cut each other.

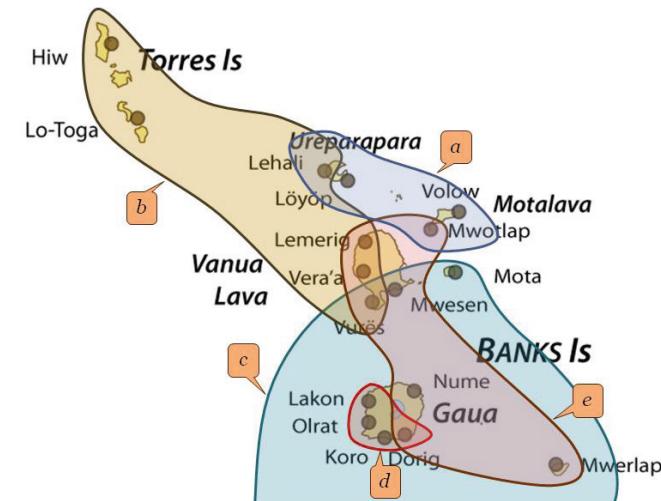
The innovations under discussion here are of various kinds (François 2011a:192–211). They include regular phonological change; irregular sound change (which affects one or a few words rather than applying across the lexicon); morphological change; syntactic change; and lexical replacement. Map 3 shows a selection of isoglosses for the following five innovations:

- a) Regular sound change: \**r* > /j/
- b) Irregular sound change: \**malate* → \**malete* ‘broken’
- c) Irregular sound change: \*?*yayaRi* → \*?*yajai* ‘almond’
- d) Morphological change: metathesis in trial pronouns  
(Plural+three → three+Plural)
- e) Morphological change: \**toya* ‘stay’ → Prohibitive

Map 3 makes it clear that isoglosses in the Torres and Banks languages—like those in the Romance family—constantly intersect.<sup>10</sup> There is no way the genealogical relations among these languages could be represented by a tree. François (2004) was an attempt to do precisely this; while a tentative tree was indeed proposed, the number of

<sup>10</sup> Note that one innovation, namely (c), involves not only a subset of the Banks languages, but also languages further south in Vanuatu (François 2011b:157).

issues raised (conflicting evidence, intersecting isoglosses, the need to constantly resort to *ad hoc* hypotheses to preserve the tree structure) were preliminary signs of the inadequacy of the cladistic approach in this part of the world.



Map 3: Five isoglosses in the Torres–Banks Islands

What we need here is a Historical Glottometry approach, which will tell us, amongst the 131,070 ( $= 2^{17} - 2$ ) potential groupings involving these languages, which ones actually exist, and constitute the strongest subgroups. That these subgroups will probably intersect is to be expected, and is no longer a problem: by now there is good reason to believe that this is the default situation in most language families. What we need is simply to go beyond the observation of individual isoglosses as in Map 3, and be able to base our calculations on a rich database.

#### 4.3. Identifying innovations

##### 4.3.1. Applying the Comparative Method

Our dataset consists of a table of 474 separate innovations which A. François identified in these 17 languages. For each linguistic feature considered, systematic comparison was conducted among languages of the sample as well as with other Oceanic languages, following principles of the Comparative Method, so as to establish the ancestral state of each property in the languages' shared ancestor (Proto-Oceanic, or a close variant thereof) as well as the direction of change.

Some cases make it relatively easy to determine what the innovation was. For example, consider the words for ‘almond’: whereas the eight languages to the north reflect the protoform \**yayaRi* (e.g. Vera'a *yar*), the languages further south reflect a form \**yajai* (e.g. Vurës *ye*). The latter protoform shows the irregular loss of \*R, a frequent yet

lexically-specific sound change in the area (François 2011b). It is clearly an innovative form, whose distribution in the Banks Islands is represented by isogloss (c) in Map 3.

In other cases, identifying the innovation requires more reflection. For example, most of the northern Vanuatu languages have an adjective meaning ‘broken’, with forms that are cognate with each other:

- (1) ‘broken’: HIW *mijit*; LTG *məlit*; LHI *melet*; LYP *malat*; VLW *malat*; MTP *malat*; LMG *mele?*; VRA *mlit?*; VRS *mlit*; MSN *malat*; MTA *malate*; NUM *malat*; DRG *mlat*; MRL *melet*.

One can show that these modern forms go back to two distinct protoforms: *\*malate* and *\*malete*. This conclusion is based on our knowledge of regular sound changes in this area, established using the Comparative Method (François 2005). This allows us to discern even those cases where two cross-linguistic homophones derive from different etyma: for example, while Lehali /melet/ necessarily reflects *\*malete*, the same surface form /melet/ in Mwerlap is a regular reflex of *\*malate*, because a stressed /a/ followed by an unstressed /e/ in the next syllable regularly underwent umlaut in this language (*\*aCe* > /eC/). Knowledge of each language’s phonological history likewise enables us to link each modern form in (1) to one, and only one, of the two protoforms—either *\*malate* or *\*malete*. The next, crucial step consists in determining which of these two is conservative, and which one is innovative. External evidence is indispensable here, and shows that other Oceanic languages outside the Torres–Banks area point to the form with /a/: e.g. Araki /nalare/ ‘broken’ < *\*malate* (François 2002:270). In sum, the innovation we are concerned with here is a lexically-specific, irregular sound change whereby *\*malate* became *\*malete*, and not the other way around. The languages that participated in this particular innovation are: Hiw, Lo-Toga, Lehali, Lemerig, Vera'a and Vurës. This innovation is represented with isogloss (b) in Map 3.

#### 4.3.2. Creating the dataset

The sort of reasoning illustrated above, which follows a rigorous application of the Comparative Method, was used to identify all 474 innovations. The distribution of innovations into various types was as follows:

NATURE OF CHANGE	NUMBER	PROPORTION
Regular sound change	21	4 %
Irregular sound change	116	25 %
Morphological change	91	19 %
Syntactic change	10	2 %
Lexical replacement	236	50 %
<i>Total</i>	474	100 %

Among these types of changes, we consider irregular sound change and morphological change to be the most diagnostic of historical relatedness (following Greenberg 1957:51, Ross 1988:12), because they are least likely to be independently innovated. Lexical material is often excluded from subgrouping studies under the assumption that it is easily borrowable; to avoid this (perceived) problem, we have included here only those lexical replacements which can be shown to predate events of (regular or irregu-

lar) sound change.<sup>11</sup>

Figure 10 shows what the final database looks like. The 17 languages are ranked from north-west to south-east; each row corresponds to one innovation, and indicates whether there is positive evidence that a language participated (1) or did not participate (0) in the innovation. An empty box (–) was used when the data are inconclusive, non-applicable, or simply lacking. Altogether, the database contains 2728 positive (‘1’), 5040 negative (‘0’) and 290 agnostic (‘–’) data points.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	X	
115	ble	“yaRat” → “yarati	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	6	
116	ble	“yarati” → “yat”	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1		
117	scratches	“yaru” → “garu	1	1	–	–	0	0	1	1	1	1	1	1	0	0	0	0	0	0	7	
118	tex	“ye-sali”	0	0	1	–	1	1	1	1	1	1	1	1	1	1	1	1	1	1	12	
119	tex	squeezes	“yb̪lir” → “bayiri”	0	0	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	5
120	tex	unstressed “gb̪” > “kp”	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	16	
121	tex	fence	“yeora	0	0	0	0	1	1	1	0	–	0	0	0	1	1	0	0	0	0	5
122	tex	this way + Dres	“gelu”	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	5
123	tex	Reversive	“gelu”	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	10	
124	tex	“yi”	“yidara” → “darua”	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	10
125	tex	3rd suffix	“yidara” → “darua”	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8
126	tex	1st clitic	“yidara” → “yidua”	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	3
127	tex	1st clitic	“yid” > “nil” da	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	3
128	tex	“yin̪itay”	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	–	1	0	2	

Figure 9: A sample of our database of historical innovations in the Torres–Banks languages.

Note that each pattern of 1s and 0s corresponds to a diffusion area, and would be represented with an isogloss. We will now illustrate the application of Historical Glottometry to this database, following the methods explained in the previous section.

## 4.4. The results

### 4.4.1. Numerical results

The first thing we can do with this dataset is to measure cohesiveness for clusters of two languages. This measure of “pairwise cohesiveness”,<sup>12</sup> applied to all pairs of languages ( $17^2 = 289$ ), yields the results in Table 1.

The figures of 100% in the diagonal simply say, as it were, that a language always subgroups perfectly with itself; these can thus be disregarded. More instructive is the observation that the cohesiveness  $k$  of language pairs tends to vary a lot, but with the highest figure being only 92%. The coloured (yellow and orange) cells indicate rates of 50% and above, i.e. pairs with relatively high cohesiveness.

To illustrate the proper interpretation of the table, the figure of 92%, between Volow and Mwtlap, indicates that when either of these languages underwent a change (to-

<sup>11</sup> This is the same reasoning that validates *\*manducāre* ‘eat’ as a legitimate example of early lexical innovation shared by French and Italian (§2.1), because it reflects regular sound changes diagnostic of inherited vocabulary (compare French *manger* /mãʒe/ < \**manducāre* with *venger* /võʒe/ ‘avenge’ < \**vindicāre*). By contrast, a recent Italian loanword such as *caporal* (‘corporal’), which does not exhibit any such sound changes, would not normally qualify as diagnostic evidence for subgrouping.

<sup>12</sup> This is quite similar to the concept of “Relative Identity Weight” in the Salzburg school of dialectometry (Goebel 2006: 412).

gether with some other language), it shared it with the other member of the pair 92% of the time. Table 1 thus shows that languages share innovations with their immediate neighbours a lot of the time—yet they do so at varying rates.

	<i>Hiw</i>	<i>Lo-Toga</i>	<i>Lehali</i>	<i>Löyöp</i>	<i>Volow</i>	<i>Mwotlap</i>	<i>Lemerig</i>	<i>Vera'a</i>	<i>Vurés</i>	<i>Mwesen</i>	<i>Mota</i>	<i>Nume</i>	<i>Dorig</i>	<i>Koro</i>	<i>Olat</i>	<i>Lakon</i>	<i>Mwerlap</i>
<i>Hiw</i>	100	83	38	28	26	25	24	22	18	18	19	14	14	16	16	14	15
<i>Lo-Toga</i>	83	100	41	30	28	27	29	28	20	22	21	16	14	17	19	17	17
<i>Lehali</i>	38	41	100	71	63	58	53	45	33	33	34	22	19	21	24	22	24
<i>Löyöp</i>	28	30	71	100	73	71	56	47	33	35	33	20	19	20	22	20	24
<i>Volow</i>	26	28	63	73	100	92	51	42	32	33	36	23	20	21	22	20	26
<i>Mwotlap</i>	25	27	58	71	92	100	50	42	32	34	35	23	20	20	21	20	26
<i>Lemerig</i>	24	29	53	56	51	50	100	75	47	47	41	25	22	22	25	23	26
<i>Vera'a</i>	22	28	45	47	42	42	75	100	58	55	44	30	27	27	28	28	30
<i>Vurés</i>	18	20	33	33	32	32	47	58	100	85	60	45	39	34	34	31	38
<i>Mwesen</i>	18	22	33	35	33	34	47	55	85	100	61	44	39	34	37	33	40
<i>Mota</i>	19	21	34	33	36	35	41	44	60	61	100	58	45	38	37	34	52
<i>Nume</i>	14	16	22	20	23	23	25	30	45	44	58	100	71	57	50	46	65
<i>Dorig</i>	14	14	19	19	20	20	22	27	39	39	45	71	100	78	64	57	59
<i>Koro</i>	16	17	21	20	21	20	22	27	34	34	38	57	78	100	82	73	50
<i>Olat</i>	16	19	24	22	22	21	25	28	34	37	37	50	64	82	100	89	47
<i>Lakon</i>	14	17	22	20	20	23	28	31	33	34	46	57	73	89	100	44	44
<i>Mwerlap</i>	15	17	24	24	26	26	26	30	38	40	52	65	59	50	47	44	100

**Table 1:** Pairwise cohesiveness values (percentages) among the 17 Torres–Banks languages

These figures, incidentally, are a valuable result in themselves, as they provide an empirical measurement of how much two languages have evolved together throughout their history. For example, the fact that Lo-Toga (#2) and Lehali (#3) shared only 41% of their innovations together points to a rather strong social divide between the Torres islands on the one hand, and the Banks islands on the other hand: clearly, the Lo-Toga community has had much less social interaction with Lehali ( $k = 41\%$ ) than with Hiw ( $k = 83\%$ ). Likewise, it is instructive to observe that, even though the language Vurés is geographically spoken only a couple of hours' walk from Vera'a (see Map 2), the two languages share together no more than 58% of their innovations; the historical links were much stronger, on the one hand, between Vera'a and Lemerig ( $k = 75\%$ ), and on the other hand, between Vurés and Mwesen ( $k = 85\%$ ). Interestingly, these figures closely match the intuitive feel one gets when learning and comparing the languages of Vanua Lava, as well as the islanders' own impressions; except that the figures have the advantage of being precise, and directly comparable with one another.

In order to deserve the status of genealogical subgroup, a cluster of languages needs to be “attested” historically, i.e. have at least one exclusively shared innovation ( $\epsilon \geq 1$ ). A subgroup uniting Volow and Löyöp, for example, would have high cohesiveness (73%) if it existed; but because no innovation happens to be shared exclusively by these two languages, they cannot count together as a subgroup. Pairings that are not supported by at least one isogloss appear here in orange. Conversely, the yellow cells in Table 1 correspond to those higher-cohesiveness pairings ( $k \geq 50\%$ ) which are actually attested as subgroups: e.g. Hiw–Lo-Toga with 83%, Lehali–Löyöp with 71%, etc.

We applied the same method to calculate the cohesiveness ( $k$ ) of all attested clusters of North Vanuatu, of any size. In total, the number of unique innovation-defined sub-

groups was 143. This figure includes the 15 pairs of languages shown in yellow in Table 1 above, but also clusters of various sizes, up to 15 members. The results, which cannot all be presented here for lack of space, were useful for the next stage: the calculation of subgroupiness values ( $\varsigma$ ).

#### 4.4.2. A glottometric diagram

We calculated the subgroupiness of all 143 attested language clusters, by applying the principles exposed in §3 above. The 15 subgroups with the highest subgroupiness values are listed in Table 2.

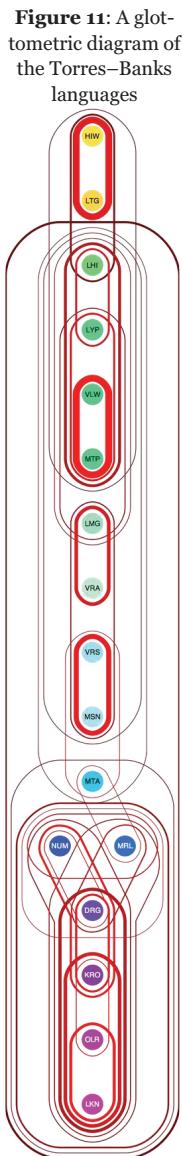
**Table 2:** The 15 strongest subgroups in the Torres–Banks linkage.

subgroups	subgroupiness
Volow–Mwotlap	12.82
Hiw–LoToga	12.45
Vurés–Mwesen	9.34
Lemerig–Vera'a	6.78
Koro–Olrat–Lakon	6.63
Dorig–Koro–Olrat–Lakon	6.01
Olrat–Lakon	5.34
Lehali–Löyöp–Mwotlap–Volow	5.22
15 Banks languages (LHI→LKN)	3.92
Dorig–Koro	3.90
Löyöp–Volow–Mwotlap	3.64
Lehali–Löyöp	3.53
Hiw–LoToga–Lehali	3.43
southern Banks (Mwerlap + Gaua)	2.99
Dorig–Mwerlap	2.37

In terms of visual representation, the abundance of subgroups of varying strengths made it necessary to represent only the strongest ones—we chose to show only those whose subgroupiness value is greater than or equal to 1 ( $\varsigma \geq 1$ ). This includes the 15 subgroups listed in Table 2, plus 17 others. We then represented each subgroup's strength by having line thickness proportional to its subgroupiness. In addition, the degree of redness (brightness value of the contour line) was made proportional to its *cohesiveness*, with more cohesive subgroups appearing more intensely red. The final result was a comprehensive **glottometric diagram** of the whole region (Figure 11).

This result would warrant more commentary than is possible in this paper;<sup>13</sup> we will stick to the essentials. First of all, the subgroupiness values, as well as the map derived from them, confirm the statement in §4.2, that the languages of northern Vanuatu form a *linkage* in which isoglosses, and hence subgroups, constantly intersect. Lehali (LHI), for example, subgroups both with the two Torres languages to its north ( $\varsigma = 3.43$ ) and with the other Banks languages to its south ( $\varsigma = 3.92$ ). Similarly Mota (MTA) forms the bridge, as it were, between a northern Banks subgroup (running from Lehali to Mota,  $\varsigma = 1.03$ ) and a distinct southern Banks subgroup (running from Mota to Lakon,  $\varsigma = 1.30$ ). No family tree could ever account for this situation.

<sup>13</sup> The colors of the dots representing the languages are also significant, but explaining this would be a task best left to a lengthier paper (François & Kalyan forthc.).



It is worthy of notice that the glottometric approach can also detect and represent those situations which are “tree-like”: for example, Volow and Mwotlap form a subgroup clearly separated from Löyöp; Vurès and Mwesen also clearly belong together. But evidently, these tree-like patches are a rarity in a language network which is strongly non-tree-like.

Another important result is the observation that Torres–Banks languages generally pattern in a geographically coherent way: all languages adjacent on the glottometric diagram are also adjacent geographically (though not vice versa; see below). This is even true for the non-linear part of the map, involving the four languages Mota–Name–Dorig–Mwerlap: all the language pairs attested there (MTA–MRL, NUM–MRL, NUM–DRG, MRL–DRG) correspond to adjacent languages on Map 2. It is impossible to capture such tight geographical organisation using a tree: any binary tree of 17 languages will allow 65,536 ( $= 2^{16}$ ) possible linear orderings of languages.

Expected though it may be, this consistency between language history and geography is a valuable result: for it shows that the languages’ anchoring in space must have remained stable over the three millennia of their historical development, with limited interisland migration (François 2011b: 181). Applying Glottometry to historically more turbulent families would make it possible to detect the genealogical relations that hold between languages *in spite* of their geographic locations, as accurately as the Comparative Method on which it is based.

And indeed, a finer grain of observation reveals certain non-trivial patterns in our data that do more than just index geography. For example, even though Volow’s location is closer to Mota than to Löyöp (Map 2), the position of the three languages in the diagram shows that Volow and Mota are genealogically quite remote ( $k = 36\%$ ). Evidently, the ancient societies of Motalava and Mota islands had very few direct social interactions with each other, and much more with the other islands (Urepidapara, Vanua Lava) located to their west. Such a result illustrates the potential of the method to reconstruct the shape of past social networks.

## 5. Conclusion

In conclusion, our newly proposed method of Historical Glottometry allows us to escape the false dichotomies of the tree model—e.g. whether it is A–B, A–C, or B–C that is truly diagnostic of genealogical relatedness—by allowing us to posit intersecting subgroups, and to quantify the strength of the genealogical evidence in favour of each language cluster.

If we were to use a tree to represent our data, we would certainly be able to capture certain salient organising features, e.g. the split between the two Torres languages (Hiw and Lo-Toga) and all the languages to the south. But a tree would only be able to provide a very distorted picture of the social history of the region—as an orderly sequence of migrations with loss of contact—while the story told by the data (made visible to us by the glottometric diagram) is a much richer and more varied narrative of social interaction in which languages converge as much as they diverge. Far from the approximations imposed by the assumptions of the tree model, we hope to have shown the way towards a more accurate and realistic representation, which stays true to the most valuable insights of the Comparative Method.

## References

- Aikhenvald, Alexandra & R. M. W. Dixon, eds. (2001). *Areal diffusion and genetic inheritance: problems in comparative linguistics*. Linguistics. Oxford: Oxford University Press.
- Alkire, Ti & Rosen, Carol. (2010). *Romance Languages: A Historical Introduction*. Cambridge: Cambridge University Press.
- Anttila, Raimo. (1989). *Historical and Comparative Linguistics*. (2<sup>nd</sup> edition). Amsterdam: John Benjamins.
- Avise, John C. (2000). *Phylogeography: The history and formation of species*. Cambridge, MA: Harvard University Press.
- Berger, Roger & Annette Brasseur. (2004). *Les Séquences de sainte Eulalie*. Geneva: Droz.
- Biggs, Bruce. (1965). Direct and indirect inheritance in Rotuman. *Lingua* 14:383–415.
- Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.
- Bossong, Georg. (2009). Divergence, convergence, contact. Challenges for the genealogical classification of languages. In *Convergence and divergence in language contact situations*, ed. by K. Braumüller & J. House. Hamburg Studies on Multilingualism, 8. Amsterdam-Philadelphia: John Benjamins. Pp.13–40.
- Bryant, David, Flavia Filimon and Russell D. Gray. (2005). Untangling our past: Languages, Trees, Splits and Networks. In *The Evolution of Cultural Diversity: Phylogenetic Approaches*, ed. by R. Mace, C. Holden & S. Shennan. London: UCL Press. pp. 69–85.
- Campbell, Lyle. (2004). *Historical linguistics: an introduction*. MIT Press.
- Carrington, Peter J., John Scott and Stanley Wasserman (eds.). (2005). *Models and Methods in Social Network Analysis*. New York: Cambridge University Press.
- Chambers, Jack K., & Peter Trudgill. (1998). *Dialectology*. Cambridge textbooks in linguistics. Cambridge: Cambridge University Press.
- Chappell, Hilary. (2001). Language contact and areal diffusion in Sinitic languages. In Aikhenvald, A. and R.M.W. Dixon (eds.), pp. 328–357.
- Croft, William. (2000). *Explaining language change: An evolutionary approach*. Pearson Education.
- Crowley, Terry & Claire Bowern. (2010). *An Introduction to Historical Linguistics*. (4<sup>th</sup> edition). Oxford: Oxford University Press.
- van Driem, George. (2001). *Languages of the Himalayas: An ethnolinguistic handbook of the greater Himalayan region*. Leiden: Brill.
- Enfield, Nicholas. (2008). Transmission Biases in Linguistic Epidemiology. In R. Nicolai & B. Comrie (eds.), *Language Contact and the Dynamics of Language: Theory and Implications*. Thema 2. Journal of Language Contact, 299–310.
- François, Alexandre. (2002). *Araki. A disappearing language of Vanuatu*. Pacific Linguistics, 522. Canberra: Australian National University.
- François, Alexandre. (2004). Subgrouping hypotheses in North Vanuatu. Paper presented at the Sixth International Congress on Oceanic Linguistics (COOL6). University of the South Pacific, Port Vila, Vanuatu.

- François, Alexandre. (2005). Unraveling the history of the vowels of seventeen northern Vanuatu languages. *Oceanic Linguistics* 44 (2):443–504.
- François, Alexandre. (2011a). Social ecology and language history in the northern Vanuatu linkage: A tale of divergence and convergence. *Journal of Historical Linguistics* 1 (2):175–246.
- François, Alexandre. (2011b). Where \*R they all? The geography and history of \*R loss in Southern Oceanic languages. *Oceanic Linguistics* 50 (1):140–197.
- François, Alexandre. (2012). The dynamics of linguistic diversity: Egalitarian multilingualism and power imbalance among northern Vanuatu languages. *International Journal of the Sociology of Language* 214:85–110.
- François, Alexandre. (forthc.). Models of language diversification. In Claire Bowern & Bethwyn Evans (eds), *The Routledge Handbook of Historical Linguistics*. New York: Routledge.
- Garrett, Andrew. (2006). Convergence in the Formation of Indo-European Subgroups: Phylogeny and Chronology. *Phylogenetic methods and the prehistory of languages*, ed. by Peter Forster & Colin Renfrew, 139–151. Cambridge: McDonald Institute for Archaeological Research.
- Geraghty, Paul A. (1983). *The History of the Fijian languages*. Oceanic Linguistics Special Publication, 19. Honolulu: University of Hawaii Press.
- Goebel, Hans. (2006). Recent Advances in Salzburg Dialectometry. *Literary and Linguistic Computing* 21 (4): 411–435.
- Gray, Russell D., David Bryant, & Simon J. Greenhill. 2010. On the shape and fabric of human history. *Philosophical Transactions of the Royal Society London, B* 365:3923–3933.
- Greenberg, Joseph H. (1957). *Essays in Linguistics*. Chicago: University of Chicago Press.
- Hall, Robert A. Jr. (1950). The Reconstruction of Proto-Romance. *Language* 26 (1):6–27.
- Hashimoto, Mantarō J. (1992). Hakka in *Wellentheorie* perspective. *Journal of Chinese Linguistics* 20:1–49.
- Haspelmath, Martin. (2004). How hopeless is genealogical linguistics, and how advanced is areal linguistics? *Studies in Language*, 28(1), 209–223.
- Heggarty, Paul, Warren Maguire & April McMahon. (2010). Splits or Waves? Trees or Webs? How Divergence Measures and Network Analysis can Unravel Language Histories. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:3829–3843.
- Hock, Hans Henrich. (1991). *Principles of Historical Linguistics*. (2<sup>nd</sup> edition). Berlin: Mouton de Gruyter.
- Holton, Gary. 2011. A Geo-linguistic Approach to Understanding Relationships within the Athabaskan Family. Paper read at the international workshop *Language in Space: Geographic Perspectives on Language Diversity and Diachrony*. Boulder, Colorado.
- Huehnergard, John, & Aaron Rubin. (2011). Phyla and Waves: Models of Classification. Semitic Languages. In *Semitic Languages: An International Handbook*, edited by S. Weninger, G. Khan, M. P. Streck & J. Watson. Handbücher zur Sprach- und Kommunikationswissenschaft, 36. Berlin: de Gruyter Mouton. Pp.259–278.
- Kortlandt, Frederik. (2007). *Italo-Celtic origins and prehistoric development of the Irish language*. Amsterdam: Rodopi.
- Krauss, Michael E., & Victor Golla. (1981). Northern Athapaskan languages. In *Handbook of North American Indians*, vol. 6: *Subarctic*, edited by J. Helm. Pp.67–85.
- Labov, William. (1963). The social motivation of sound change. *Word* 19:273–309.
- Leskien, August. (1876). *Die Declination im Slawisch-Litauischen und Germanischen*. Leipzig: Hirzel.
- Milroy, James, & Lesley Milroy. (1985). Linguistic change, social network and speaker innovation. *Journal of linguistics* 21 (2):339–384.
- Minaka, Nobuhiro & Kunihiko Sugiyama. (2012). *Keitōju Mandara: Chein/tsuri/nettowāku* [Phylogeny Mandala: Chain, tree, and network] (in Japanese). Tokyo: NTT Publishing.
- Nerbonne, John. (2010). Measuring the diffusion of linguistic change. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:3821–3828.
- Page, Roderick D. M. & Edward C. Holmes. (2009). *Molecular Evolution: A phylogenetic approach*. Oxford: Blackwell.
- Pawley, Andrew. (1999). Chasing Rainbows: Implications of the Rapid Dispersal of Austronesian Languages for Subgrouping and Reconstruction. *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun & Paul Jen-Kuei Li, 95–138. Symposium Series of the Institute of Linguistics: Academia Sinica.
- Pawley, Andrew. (2008). Where and when was Proto-Oceanic spoken? Linguistic and archaeological evidence. In *Language and text in the Austronesian world: studies in honour of Ülo Sirk*, edited by Y. A. Lander & A. K. Oglöblin. München: Lincom Europa. Pp.47–71.
- Pawley, Andrew. (2009). Polynesian paradoxes: Subgroups, wave models and the dialect geography of Proto Polynesian. Unpublished paper delivered at the Eleventh *International Conference on Austronesian Linguistics*, Aussois (France), June 2009.
- Pawley, Andrew. (2010). Prehistoric Migration and Colonisation Processes in Oceania: A View from Historical Linguistics and Archaeology. In *Migration history in world history: multidisciplinary approaches*, edited by J. Lucassen. Studies in global social history, 3. Leiden: Brill. Pp.77–112.
- Pawley, Andrew, & Roger C. Green. (1984). The Proto-Oceanic Language Community. *Journal of Pacific History* 19:123–146.
- Pawley, Andrew, & Malcolm Ross. (1995). The prehistory of Oceanic languages: a current view. In *The Austronesians: Historical and Comparative Perspectives*, edited by P.S. Bellwood, J.J. Fox & D. Tryon, Comparative Austronesian Project. Canberra: Australian National University. Pp.39–80.
- Posner, Rebecca. (1996). *The Romance Languages*. Cambridge: Cambridge University Press.
- Ross, Malcolm. (1988). *Proto-Oceanic and the languages of Western Melanesia*. Canberra: ANU Press.
- Ross, Malcolm. (1997). Social networks and kinds of speech-community event. In *Archaeology and language 1: Theoretical and methodological orientations*, edited by R. Blench & M. Spriggs. London: Routledge. Pp.209–261.
- Ross, Malcolm, Andrew Pawley, & Meredith Osmond, eds. (2011). *The lexicon of Proto Oceanic: Animals*. Pacific Linguistics, 621. Canberra: Australian National University.
- Schmidt, Johannes. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Hermann Böhlau.
- Schmidt, Karl-Horst. (1993). Insular Celtic: P and Q Celtic. In *The Celtic languages*, edited by M. J. Ball & J. Fife. New York: Routledge. Pp.64–99.
- Séguy, Jean. (1973). La dialectométrie dans l'Atlas linguistique de la Gascogne. *Revue de linguistique romane* 145–146:1–24.
- Skelton, P., Smith, A., & Monks, N. (2002). *Cladistics: a practical primer on CD-ROM*. Cambridge University Press.
- Southworth, Franklin C. (1964). Family-tree diagrams. *Language* 40(4):557–565.
- Szemerédy, Benedikt. (2011). Corpus-based dialectometry: a methodological sketch. *Corpora* 6 (1):45–76.
- Toulmin, Matthew. 2009. *From Linguistic to Sociolinguistic Reconstruction: the Kamta historical subgroup of Indo-Aryan*. Pacific Linguistics: Australian National University.
- Tryon, Darrell. 1996. Dialect chaining and the use of geographical space. In *Arts of Vanuatu*, edited by J. Bonnemaison, K. Huffman, C. Kaufmann & D. Tryon. Bathurst: Crawford House Press. Pp.170–181.
- Valente, Thomas W. (1995). *Network Models of the Diffusion of Innovations*. Cresskill, NJ: Hampton Press.
- Wüst, Jakob. (1994). La restructuration du système des démonstratifs en protoroman, in Jacqueline Cerguiglini-Toulet and Olivier Collet (eds.), *Mélanges de philologie et de littérature médiévales offerts à Michel Burger*, 41–49. Genève: Droz.

---

**ANU College of Asia-Pacific  
School of Culture, History and Language**

***Department of Linguistics***

Bethwyn Evans  
Research Fellow, Linguistics  
School of Culture, History and Language  
College of Asia and the Pacific  
Coombs Building  
The Australian National University

Canberra ACT 0200 Australia  
T: +61 2 6125 0300  
E: [bethwyn.evans@anu.edu.au](mailto:bethwyn.evans@anu.edu.au)  
[www.anu.edu.au](http://www.anu.edu.au)

18 February 2014

TO WHOM IT MAY CONCERN

I write as one of the editors of the following volume:

Bowern, Claire and Bethwyn Evans (eds). 2014. *The Routledge Handbook of Historical Linguistics*. London: Routledge.

This volume comprises chapters representing the state-of-the-art and future directions of historical linguistics, and is currently in press. It is to be published with the academic publisher Routledge, and is with the Routledge copy-editing team at present. We will receive the final proofs in March, and the expected publication date is June 2014.

This letter is to certify that the chapter written by Dr Alexandre François, entitled *Trees, Waves and Linkages: Models of Language Diversification*, has been reviewed by the editors of the volume, and deemed suitable for inclusion in the handbook. This chapter, as with all other chapters, has been finalised and is with the Routledge copy-editing team who are preparing the final proofs.

Yours sincerely,

*Bethwyn Evans*

Bethwyn EVANS



# Trees, waves and linkages: Models of language diversification

---

Alexandre FRANÇOIS  
CNRS-LACITO — Australian National University

*Contrary to widespread belief, there is no reason to think that language diversification typically follows a tree-like pattern, consisting of a nested series of neat splits. Except for the odd case of language isolation or swift migration and dispersal, the normal situation is for language change to involve multiple events of diffusion across mutually intelligible idiolects in a network, typically distributed into conflicting isoglosses. Insofar as these events of language-internal diffusion are later reflected in descendant languages, the sort of language family they define – a “linkage” (Ross 1988) – is one in which genealogical relations cannot be represented by a tree, but only by a diagram in which subgroups intersect.*

*Non-cladistic models are thus needed to represent language genealogy. This chapter focuses on an approach that combines the precision of the Comparative Method with the realism of the Wave Model. This method, labeled Historical Glottometry, identifies genealogical subgroups in a linkage situation, and assesses their relative strengths based on the distribution of innovations among modern languages. Provided it is applied with the rigour inherent to the Comparative Method, Historical Glottometry should help unravel the genealogical structures of the world's language families, by acknowledging the role played by linguistic convergence and diffusion in the historical processes of language diversification.*

## 1 ON THE DIVERSIFICATION OF LANGUAGES

### 1.1 Language extinction, language emergence

The number of languages spoken on the planet has oscillated up and down throughout the history of mankind.<sup>1</sup> Different social factors operate in opposite ways, some resulting in the decrease of language diversity, others favouring the emergence of new languages. Thus, languages fade away and disappear when their speakers undergo some pressure towards abandoning their heritage language and replacing it in all contexts with a new language that is in some way more socially prominent (Simpson, this volume). The process of language extinction may be rapid or slow, and varies in intensity depending on historical circumstances.

While this process results in the erosion of language diversity, others bring about the opposite result: an increase in the number of spoken languages. Because no natural language appears *ex nihilo*, one has to explain how new languages emerge out of older ones. Some – such as pidgins and creoles (Romaine 1988, Siegel 2004) or mixed languages (Matras & Bakker 2003) – result historically from the encounter of two populations who were driven, under very special social conditions, to combine elements of their respective languages and create a new one. Yet this pattern, whereby a language is born of two parents, is not the typical scenario. New languages also commonly arise from the internal diversification of a single language as it evolves into separate daughter languages over time, following processes where external input does not necessarily play the central role. This phenomenon of internal diversification is the object of the present chapter.

The two tendencies outlined above – language extinction and language emergence – have always occurred in human history;<sup>2</sup> yet in terms of scientific knowledge, the modern scholar is faced here with a strong asymmetry. Except for the few that have left behind written materials that can be deciphered, most extinct languages of the past will forever be unknown, whether in their linguistic structures or the social causes of their demise. By contrast, linguistic diversification has brought about an observable outcome, in the form of attested languages. The latter can be analysed and compared in a historical perspective, thereby bringing invaluable insights into their linguistic and social development. This asymmetry in the availability of data explains why the process of *language diversification* plays such a central role in the discipline of historical linguistics. The aim of the present chapter is to understand how this process of diversification takes place in languages, and what model can best account for the empirically observed patterns of language relations.

## 1.2 Trees vs. waves: two models of language diversification

Our point of departure is the observation that several modern languages can historically stem from the internal diversification of what was once a single language, with no need to resort primarily to external factors such as contact or language admixture. The internal diversity among modern Romance languages, for example, can largely be explained by a process of internal fragmentation, taking a relatively homogeneous variety of spoken Latin as a starting point. While contact-related factors – substrate, superstrate and adstrate influences involving non-Romance languages – did play their part, a large proportion of the history of Romance can be reconstructed as internal diversification affecting inherited linguistic material.

For most language families, unlike in Romance, the ancestral language is not attested but merely hypothetical; the reconstruction of historical scenarios leading to modern languages is then the object of logical analysis and the weighing of competing hypotheses, based on a systematic comparison of the attested languages. This procedure, known as the COMPARATIVE METHOD (see chapters by Weiss and Hale in this volume), was initially developed by the German Neogrammarians in the second half of the 19th century, and

constitutes, to this day, the most successful approach in reconstructing the history of language families.

The Comparative Method has tended to be closely associated with a particular model of diversification: the *Stammbaum*, or family tree. Ever since this model was first proposed by August Schleicher in his 1853 article *Die ersten Spaltungen des indogermanischen Urvolkes*, its association with the Comparative Method has been taken for granted (e.g. Bloomfield 1933:311; Campbell 2004:165; etc.); yet I will claim here that the two lines of thinking ought to be dissociated. While the Comparative Method is without a doubt the most solid approach to the reconstruction of language histories, I will argue that the Tree Model presupposes a flawed understanding of language diversification processes. In a nutshell, cladistic (tree-based) representations are entirely based on the fiction that the main reason why new languages emerge is the abrupt division of a language community into separate social groups. Trees fail to capture the very common situation in which linguistic diversification results from the fragmentation of a language into a network of dialects which remained in contact with each other for an extended period of time (Bloomfield 1933; Croft 2000; Garrett 2006; Heggarty, Maguire & McMahon 2010; Drinka 2013), creating what Ross (1988, 1997) calls a “linkage” (see §3.3).

The present chapter will discuss the strengths and weaknesses of cladistic representations for modelling processes of language diversification, and examine alternative approaches for capturing the genealogy<sup>3</sup> of languages. In section 2, I will first summarise the way in which linguistic trees are typically understood, before examining their underlying assumptions. Section 3 will examine the processes that underlie genealogical relations between languages, and explain why the Tree Model is most often unsuited for representing them. While the Comparative Method must be preserved for its invaluable scientific power, a rigorous application of its principles in situations of linkage in fact disproves the Tree Model, and favours the WAVE MODEL (§3.2) as a more accurate description of the genealogy of languages.

Non-cladistic models are needed to represent language relationships, in ways that take into account the common case of linkages and intersecting subgroups. Among existing models, Section 4 will focus on an approach that combines the precision of the Comparative Method with the realism of the Wave Model. This method, labeled *Historical Glottometry* (Kalyan & François f/c), identifies genealogical subgroups in a linkage situation, and assesses their relative strengths based on the distribution of innovations among modern languages. Provided it is applied with the rigour inherent to the Comparative Method, Historical Glottometry should help unravel the genealogical structures of the world’s language families, by acknowledging the role played by linguistic convergence and diffusion in the historical processes of language diversification.

## 2 UNDERSTANDING THE TREE MODEL

### 2.1 Reading and drawing language trees

I first propose to examine how language trees are classically understood. Let there be five modern languages, labelled K, L, M, N, O. These languages are believed to be genealogically related if they comply with a number of conditions (Campbell & Poser 2008: 162 sqq.): in particular, a sizeable number of demonstrably cognate items in their morphology and basic vocabulary, displaying regular sound correspondences in ways that cannot be reasonably assigned to chance or borrowing (Weiss, this volume).

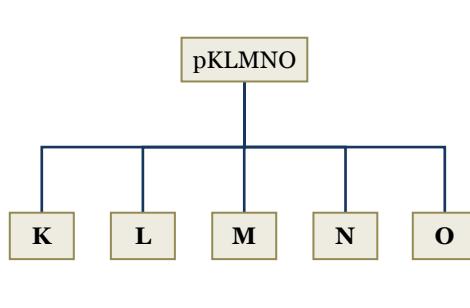


Figure 1 – An unordered genealogical tree

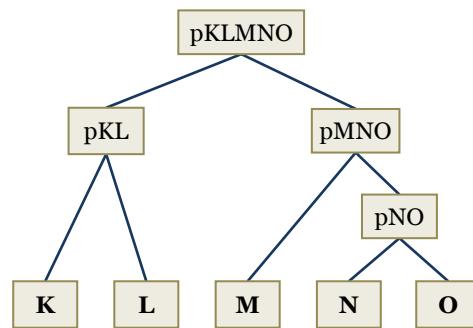


Figure 2 – A genealogical tree indicating internal subgrouping

To say that K, L, M, N, O are genealogically related entails that they ultimately descend from a common ancestor – a “proto-language”, which in this case can be called Proto-KLMNO. This point could be shown using *Figure 1*, a “rake-like” or “fan-like” representation: this shows each language as an independent descendant of the protolanguage, with no claim about the family’s internal structure. Such a “flat” tree may sometimes correspond to an actual historical situation, as when an ancestral society swiftly broke up into a number of separate subcommunities, quickly followed by a loss of mutual social contact; according to Pawley (1999), this scenario may indeed have characterised the breakup of Proto-Oceanic into lower-level subgroups. In other cases, a representation like *Figure 1* simply reflects a linguist’s agnostic view of a family’s internal structure, for instance due to lack of sufficient data. What historical linguists typically hope to achieve with a tree is to identify a number of internal subgroups within the family, into which languages with more recent shared ancestors can be grouped together. *Figure 2* illustrates the sort of ideal tree aimed at by subgrouping studies.

Such a tree captures a set of claims about the internal structure of a language family. Here, a claim is made that languages K and L “subgroup” together, by contrast with M, N and O which form their own subgroup MNO; within the latter, a claim is made that N and O form a subgroup of their own apart from M. Following a nested pattern, the language N is said to belong to the NO subgroup, which in turn forms a “branch” of the larger subgroup MNO. Even though such claims about the internal structure of a family could be formulated, in principle, in purely taxonomic terms with no reference to time, it is

common practice to interpret such cladistic representations of language families in historical terms. A common assumption is that the sequence of nodes in a tree, from top to bottom, mirrors the actual chronological order of historical events. Another frequent, and somewhat simplistic, conception (as underlined by Pulgram 1961) is that each node in the tree corresponds to an individual language community, so that a split in a tree can essentially be equated with the division of an earlier unified community into separate social groups.

Thus, to say that M, N and O subgroup together as opposed to other languages of their family, amounts to claiming that they all descend from an intermediate protolanguage – call it Proto-MNO – that was once spoken by a single social community, after the breakup of the earlier language Proto-KLMNO. According to *Figure 2*, this language Proto-MNO must have developed more or less separately from Proto-KL, the shared ancestor of modern languages K and L. This point is established through the identification of a number of linguistic innovations of various sorts (phonological, grammatical, lexical, etc.) which are jointly reflected by modern languages M, N and O, but not by other languages of the family. If these three languages share together certain linguistic properties that were not inherited from their ultimate ancestor, it is assumed – provided one can rule out chance similarity or parallel innovation – that they must have acquired these properties at a certain point in time, when their speakers still spoke (mutually intelligible variants of) a single language. The idea is that, instead of positing the same change in three languages (M, N, O) independently, it is more parsimonious – following Occam’s razor – to propose that it took place just once in a single language (Proto-MNO) and then was simply inherited by its descendants. By contrast, the fact that K and L do not reflect those innovations suggests that their ancestors did not participate in that Proto-MNO speech community. This scenario is visually summarised by the existence of the “MNO” node in *Figure 2*.

Following a principle first formulated by Leskien (1876), the Comparative Method establishes the existence of every intermediate node in a family tree based on the principle of *exclusively shared innovations*, i.e. by identifying those linguistic changes that are shared by all of its modern descendants, and only by them – what phylogeneticists call *synapomorphies* (Page & Holmes 2009). These innovations are thought to have been introduced historically during the lifetime of the intermediate protolanguage – after the split from a higher node, and before the new split into lower nodes. The reasoning is recursive: *Figure 2* also represents the claim that the ancestors of modern speakers of M, after undergoing developments that are also reflected in N and O, at some point in time started developing independently; by contrast, the remaining ancestors of N and O kept sharing innovations for some time, until they too eventually separated.

In sum, the history of the family illustrated in *Figure 2* would be summarised by stating that what used to be a single language (pKLMNO) first split into two separate languages (pKL and pMNO), which in turn were to split again. This series of recursive splits and the resulting divergence is one possible way to understand the process of language diversification, and the emergence of new languages.

## 2.2 The tree, a model based exclusively on separation

In the classical understanding of family trees, each node is thus supposed to correspond to a specific social community that developed separately from other nodes (Fox 1995:123). The sort of separation referred to here is typically understood as an actual event of social split such as migration, whereby a previously unified society broke up into two separate communities with loss of contact. Other cases are possible, such as social isolation due to the intrusion of other languages; or the *in situ* breakup of earlier networks of communication, as communities stayed in place yet decreased their mutual contact as they began – for whatever reason – to isolate themselves from each other.<sup>4</sup>

In order to yield a robust tree-like structure like the one in *Figure 2* with intermediate nodes (as opposed to the flat structure of *Figure 1*), the process of social split must be repeated recursively across the centuries; each event of separation must have been followed by a period of stability – at least a few generations – during which innovations had the time to form and settle within the new community (Pawley & Ross 1995), before another split took place again.

This focus on divergence is both a strength and a weakness of the Tree Model. A strength, because it means that trees can help reconstruct events of social disruption when they indeed took place, and can represent them using a visually straightforward diagram. But it is also a weakness, because it distorts the reality of language diversification by shoehorning it into a one-size-fits-all, simplistic model which forces us to reconstruct events of social separation even when they never really happened, at the expense of all other possible scenarios.

Let us imagine, for the sake of discussion, that there existed a language family in the world whose development did indeed take the form of social splits, repeated over and over through the centuries of its history: such a hypothetical language family could indeed be portrayed accurately by a tree such as *Figure 2* above. In reality, no population in the world can reasonably have its history reduced to just a series of social splits with loss of contact – the scenario favoured by the Tree Model. While some families did go through such events several times in their history, in the form of successive bouts of migration or similar disruptions, these events of split, correlated with neat patterns of linguistic divergence, are always interspersed with other forms of social interaction whose linguistic impact – as we'll see below – is not compatible with a tree representation.

## 2.3 Dealing with problems in a tree structure

In the interest of the forthcoming discussion, it is important to highlight the fact that, under the Tree Model, a given language may belong to only one higher-level subgroup at a time. If M is a member of the MNO subgroup, then it cannot also be a member of a KLM subgroup at the same time: subgroups are mutually exclusive, and never intersect. This seems a sensible idea if the splits in the tree are meant to represent physical separation with no return: if the communities of pKL and pMNO were indeed separated with

complete loss of contact, then it is difficult to imagine how some modern descendants of pMNO, but not others, could share anything with pKL. This principle of separate development is central to the whole logic of subgrouping studies under a cladistic approach, and has important consequences.

Let's assume that the tree in *Figure 2* rests on sufficiently solid evidence to be deemed valid. Then let's imagine that a linguistic property is found to be shared by languages L and M, and only these two languages. This will be a problem under the Tree Model, one that will require specific hypotheses in order to account for this shared property, and still save the tree structure. For example, the shared property may be proposed to be in fact a case of *shared retention* (also known as *symplesiomorphy* in phylogenetics) from the Proto-KLMNO ancestor, a property lost by other languages (K, NO): in this case, the property would not indicate any significant genealogical link between L and M – other than their remote relatedness. Alternatively, one could argue that the property is indeed innovative, yet happened independently in L and M, whether by drift or parallel innovation (*homoplasy*).

Finally, a third hypothesis would be that the property was innovated internally in only one language, say L, and then was borrowed by the other language M via *contact* between L and M, once they had already been formed as separate languages. Even though contact between languages – also known as “horizontal transmission” or “areal diffusion” – is known to be a powerful force of language change (Lucas, this volume), it is not meant to be represented on a tree. Contact-induced change, which can take place between any two languages regardless of their relatedness, is generally considered to be a separate phenomenon from the sort of “internal change” that underlies genealogical relations. The argument is that, for a property to be borrowed between two separate languages L and M, the two languages need to already exist independently; strictly speaking, the study of their genealogy is interested in how these languages came into existence, not in what happened to them later. Thus, the many words borrowed by English from Scandinavian languages during the Viking invasions, or later from French, are not considered to form part of its genealogical makeup: the English language had by that time already acquired independent existence, as it were, as a member of the Anglo-Frisian branch of the West Germanic subgroup. Following this principle, in a tree such as *Figure 2*, a property borrowed by M from L after their separation would not be considered evidence for a genealogical subgroup LM; it would be described as an effect of contact, and understood as irrelevant for subgrouping purposes.

Several authors have expressed frustration at the Tree Model, saying that trees exclusively represent language divergence, and fail to take into account contact-induced change, or *convergence*, when reconstructing language history (e.g. Fox 1995:124; Dixon 1997; Aikhenvald & Dixon 2001; Bossong 2009; Drinka 2013). They argue that loanwords, borrowed structures and other facts of cross-linguistic diffusion form part of the linguistic history of languages as much as the material directly inherited. While the latter point is undoubtedly true, proponents of the Tree Model reply to this objection by acknowledging

that trees are only intended to capture a portion of the history of languages, namely their *genealogy* strictly speaking, and nothing more. As for other facts of language development – notably the effects of contact – they are, or at least should be, treated by other models (Campbell & Poser 2008:327). This is a valid point, which bears keeping in mind every time family trees are cited: language genealogy only forms a portion of the historical picture, and trees should not be assigned more explanatory value than they actually have.

In the following sections, the argument I will put forward against the Tree Model is reminiscent of the objection just mentioned, yet distinct from it. Let us grant that contact *between separate languages* (e.g. Old English and Old French) does not form part of their genealogical makeup, and that the model we want to design is meant to focus on the latter. My main proposal will be that trees not only omit representing language contact (which is fair enough, if it is not their objective to do so) but also, more problematically, that they even fail to accurately represent language genealogy. My argument will also be based on the problem of horizontal diffusion; yet instead of concerning facts of CROSS-LINGUISTIC DIFFUSION (contact between already separated languages), my central problem will be processes of LANGUAGE-INTERNAL DIFFUSION – i.e. the diffusion of innovations across mutually intelligible idiolects in a single language community.<sup>5</sup>

For example, the whole reasoning above about a property shared between L and M would have to be quite different if the KL and MNO clusters were never in fact physically separated, but were simply sets of dialects within a larger KLMNO group of mutually intelligible varieties still in constant contact. While it may be the case that dialects K-L have shared together one set of innovations and M-N-O another one, it is perfectly plausible that dialects L and M could *also* undergo their own set of shared innovations, during the same historical period. This is how the process of language-internal diffusion, the ultimate source of genealogical relations in languages (§3.1), can give birth to subgroups that crosscut each other: K-L; L-M; M-N-O... Such a dialect-chain situation, and more generally dialect continua and linkages (§3.3), form the Achilles' heel of the Tree Model, and are best described using a non-cladistic approach (Gray, Bryant & Greenhill 2010:3229). This issue is the focus of the next section.

### 3 THE WAVES OF DIFFUSION AT THE SOURCE OF LANGUAGE GENEALOGY

#### 3.1 Theoretical principles: genealogy reflects diffusion

Recent progress made on the sociolinguistic underpinnings of language change provides an opportunity to rethink the process of linguistic diversification, and to redefine what we mean by ‘genealogical’ or ‘genetic’ relations in languages. In particular, one assumption held by the founders of the Tree Model was that the normal locus of linguistic innovations is a ‘language’ or a ‘proto-language’, understood as a monolithic unit that could be represented as a simple node in a tree. Thus for modern languages M, N and O to share the same innovation *i* would be interpreted as evidence that these necessarily descended from a single *language* (labelled Proto-MNO). Positing such a node in the tree makes it

then possible to state that the innovation *i* happened just “once” in that single language – with the assumption that this would be more parsimonious than positing parallel innovation or late contact between three separate languages M, N, O (§2.1). The whole design of the family tree rests on this fiction that a “language” unproblematically forms an atomic unit, and that innovations just “happen” in them.

This simplistic view was challenged as early as the end of the 19th century by the work of dialectologists (Gilliéron 1880, Wenker 1881), who showed that a given language typically consists of a network of dialects that can show a great deal of diversity. Language properties were found to be distributed in space following complex patterns, described visually using *isoglosses*. Far from always coinciding neatly, the default situation for these isoglosses is to target different segments of the social network, and thus intersect (cf. Trudgill 1986, Chambers & Trudgill 1998; Fox 1995:129). These views from dialectology were enriched by sociolinguistic studies, which observed how individual instances of language change are reproduced and diffused by speakers in their daily communication (see Labov 1963, 1994, 2001, 2007; Milroy 1987; Milroy & Milroy 1985). These works emphasised not only the complex geographical distribution of properties, but the intricate patterns whereby tokens of innovative features are statistically distributed in the speech of individuals, depending on a variety of social factors.

When approaching language change, the proper operational unit of observation is not the language or the dialect, but the IDIOLECT, i.e. the linguistic competence of an individual speaker at a certain point in time.<sup>6</sup> As for dialects and languages, they form more or less homogeneous systems shared by a network of *mutually intelligible idiolects*. When historical linguists identify a change that happened “once” in a “language”, they really encapsulate a long process of diffusion that took place across large networks of idiolects, sometimes spanning across several generations.

Indeed, linguistic innovations first emerge in the speech of certain individuals, in the form of novel ways of speaking – whether phonetic, lexical, phraseological, etc. If that innovation presents some sort of appeal to the hearer as a way to potentially increase their communicative goals, they may adopt it into their own speech, and start replicating it in new situations. If carried out repeatedly and extensively across a social network, this process of imitation or “accommodation” (Street & Giles 1982; Trudgill 1986; Giles & Ogay 2007) results in the spread of a new speech habit from one person to the other, across idiolects – a phenomenon which has been labelled *propagation* (Croft 2000) or linguistic *epidemiology* (Enfield 2003, 2008). After a period of competition with the previous norm, the innovation may become statistically dominant, and settle in the speech habits of a whole social group. If it does, then it becomes a property of an entire “communalect” (i.e. sociolect, dialect or language). From that point onwards, the linguistic feature will be transmitted to descendant generations of learners, just as much as the rest of the inherited system.

This language-internal diffusion of innovations gives rise to the genealogical relations among languages which subgrouping studies precisely seek to unravel. Such a process is

not fundamentally different from what is involved in language contact: both forms of diffusion involve the progression of a new linguistic behaviour across a social network of individual speakers – a process that is not reducible to a single event. The main distinction is that *contact* is normally a process of diffusion observed across separate languages, whereas *language-internal diffusion* involves mutually intelligible idiolects, which together may be taken to form a single (more or less homogeneous) language community.<sup>7</sup>

An innovation diffusing through a community may eventually (sometimes after several generations) settle into the mainstream usage of an entire network of mutually-intelligible idiolects, and thus become a feature of “the language”. When this happens, the change may be captured using a synthetic formula of the type  $x > y$ ; it may even be understood, in retrospect, as though it were a single event that took place “once” in that “language”. However, the patterns of propagation are often more complex. Specifically, the language-internal diffusion of innovations does not have to target an entire language community, and commonly settles down to just a cluster of dialects, so that successive innovations target different segments of the network. In this case, the intricate patterns resulting from language-internal diffusion cannot be captured by a tree, and need to be accounted for by a different model.

### 3.2 The Wave Model

Just such a line of theoretical reasoning underlies the “Wave Model”, or *Wellentheorie*, which Hugo Schuchardt and Johannes Schmidt proposed in the early 1870s (Schmidt 1872), as an alternative to August Schleicher’s Tree Model (*Stammbaumtheorie*). These authors occasionally conceived their Wave Model as a challenge not only to the Tree Model, but to the Comparative Method as a whole: Schuchardt, for example, linked it with a general disbelief in the Neogrammarians’ views on the regularity of sound change (Schuchardt 1885). Such an extreme stance is however not essential to the Wave Model, and unduly throws the baby (the Comparative Method) out with the bathwater (the Tree Model). A synthesis should be possible, which preserves the principle of regularity and other useful tenets of the Comparative Method, yet replaces the simplistic tree representations with a wave-inspired approach.

Under the Wave Model, each instance of language change arises somewhere within the network, and from there diffuses to adjacent speaker groups. The propagation of the change can thus be compared to a “wave” which expands away from its centre as the new feature is adopted across a broader territory. These waves are independent of each other, and are not necessarily nested. As Bloomfield (1933:317) puts it, “[d]ifferent linguistic changes may spread, like waves, over a speech-area, and each change may be carried out over a part of the area that does not coincide with the part covered by an earlier change”. Likewise, an innovation targeting a small cluster of dialects can be followed by a later one targeting a larger group.<sup>8</sup> Both these patterns are incompatible with a tree.

I will illustrate this point first with a general model, before mentioning actual examples. Each event of language change defines its own isogloss, i.e. a (typically) geographically

contiguous zone, representable on a map, within which the innovation diffused across idiolects and settled. In a linguistic continuum characterised by mutual intelligibility across adjacent dialects, the normal situation is for these isoglosses to intersect constantly, rather than be nested. For instance, let there be eight close dialects labelled A to H, and six innovations numbered #1 to #6, such that innovation #1 arose in dialect D and spread to adjacent dialects until it covered the zone CDE; #2 encompassed AB; #3 spanned over CDEF; #4 over FG; #5 over EF, and #6 over EFGH (Figure 3).

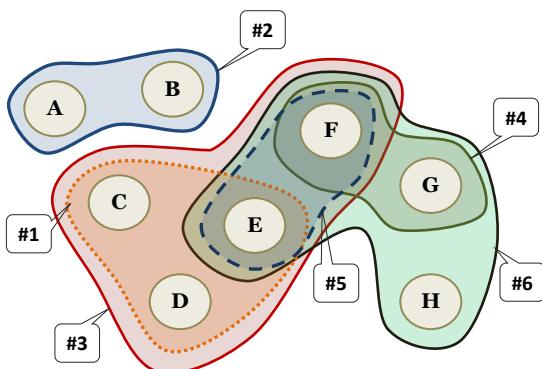


Figure 3 – Intersecting isoglosses in a dialect continuum or a linkage

The first innovations which targeted, say, the dialects C-D-E, were not radical enough to prevent mutual intelligibility with the other dialects: in the absence of a physical boundary between them, nothing then prevented the next innovation from targeting a cluster E-F, then F-G, etc. In this model, every innovation constitutes an instance of linguistic *convergence* – for the dialects that participate together in that innovation, e.g. E and F in #5 – as much as it is a case of linguistic *divergence* – for the dialects that become differentiated as a result of the change, e.g. E and D in #5 (cf. François 2011a:231).

Over time, the layered innovations leave their footprint in each local dialect. Consider a pair of dialects, for example F and G. While the changes they share together (#4, #6) have increased their similarity in certain aspects of their systems, those which have affected only one of them (either alone, or together with other neighbouring dialects – e.g. #3, #5) have increased their difference. Should many more crosscutting innovations (or “non-shared innovations”) accumulate over generations, what started as mutually intelligible dialects F and G will become opaque to each other, and eventually become distinct languages. Unless later processes of dialect levelling (or *koineisation*) take place,<sup>9</sup> each member of the network will inherit in its local system the innovations it has participated in, and these will be transmitted to its descendants. In this regard, all the innovations mentioned here, and represented in Figure 3, define the genealogical structure of the family.

As these dialects increase their differences and lose mutual intelligibility, the end result is an increase in the number of distinct languages. Yet crucially, whereas the Tree Model assigns linguistic diversification to social splits with loss of contact (§2.2), the Wave Model is compatible with scenarios where communities remain in contact. In fact, it treats

linguistic contact – in the form of multiple, criss-crossing events of diffusion across mutually intelligible dialects – as the very key to understanding patterns of language diversification. This is a radical shift in perspective.

An important implication of the Wave Model is that a given language can perfectly well belong to several partially overlapping subgroups. A GENEALOGICAL SUBGROUP is here defined as a group of languages whose ancestors participated together in the diffusion of one or several linguistic innovations, at a time when they were mutually intelligible. Crucially, nothing in this definition entails that subgroups should be discrete or nested, and indeed my claim is that genealogical subgroups can perfectly intersect, and commonly do. Thus in *Figure 3*, it is legitimate to say that E belongs simultaneously to the subgroups CDE, EF, CDEF, and EFGH – a situation which no orthodox family tree would ever be able to represent (§2.3).

### 3.3 From dialect continua to linkages

The issue of isogloss intersection has long been central to dialect geography (see Bloomfield 1933:321). It thus comes as no surprise that dialectologists, who observe the fine-grained distribution of linguistic features in space, tend to favour the Wave Model – or some model derived from it – over cladistic representations. The networks of Italian, Dutch or Arabic dialects, to take just a few examples, could never be modelled by any tree.

One could propose that the two models are complementary, in the sense that trees would be well-designed to represent the genealogical relations between separate LANGUAGES; whereas waves would only be concerned with the complex relations between DIALECTS *within* the boundaries of each language. The two models would then both be useful, but at different grains of observation. I think this view is wrong, for one important reason: namely, that many language (sub)families – as we will see below – have in fact arisen from the diversification of former dialect continua. To the extent that earlier local innovations are faithfully transmitted across generations,<sup>10</sup> the resulting languages normally keep the traces of their entangled isoglosses. If trees fail to represent genealogical relations between dialects, then they must also fail to capture the relations between the languages that descend from them.

This important point has been made by Malcolm Ross, around the concept of *linkage* (Ross 1988, 1996, 1997, 2001). Ross (1988:8) defines a linkage as “a group of communalects which have arisen by dialect differentiation”, where ‘communalect’ is a generic term which may refer to modern dialects or languages. When a dialect continuum – typically structured along the lines of *Figure 3* above – evolves in such a way that its members lose mutual intelligibility, it becomes a linkage. A linkage thus consists of separate modern languages which are all related and linked together by intersecting layers of innovations; it is a language family whose internal genealogy cannot be represented by any tree.

While Ross initially developed this concept for the historical reconstruction of Western Oceanic languages, it clearly has wider theoretical significance. Many language families or subfamilies have been shown to be linkages – whether the authors have used that term or

not. The Oceanic languages of Fiji (Geraghty 1983), Polynesian languages (Gray, Bryant & Greenhill 2010), the Indo-Aryan languages of the Kamta region of India (Toulmin 2006, 2009), the Karnic subgroup of Pama-Nyungan (Bowern 2006), northern Athabaskan languages (Krauss & Golla 1973, Holton 2011), some parts of the Semitic family (Huehnergard & Rubin 2011), Sinitic languages (Hashimoto 1992, Chappell 2001), Western Romance (Penny 2000:9–74; Ernst *et al.* 2009), Germanic (Ramat 1998), and even Indo-European as a whole (Bloomfield 1933:316; Anttila 1985:305; Garrett 2006; Drinka 2013): these are all examples, among many others throughout the world, of language families which have been shown to result from a long history of layered innovations with entangled patterns of distribution, akin to *Figure 3* above; none of them could be accurately represented by a tree. Section 4.3 below will briefly examine a particular linkage from northern Vanuatu, and propose a way to model such linkages.

### 3.4 The tree, a special case of a linkage

Based on empirical observation of the world’s language families – as illustrated by the scholarly works cited above, and many others – it thus seems that genuinely “tree-like” families are much rarer than is usually acknowledged. This is so true, that one may question the usefulness of the Tree Model as a suitable approach for representing language genealogy altogether.

One might perhaps propose to salvage the Tree Model as a useful approximation, at least for those (sub)families which are mostly compatible with it. This would go along with the conventional wisdom that the Tree and the Wave models complement each other, and should both be preserved (Hock 1991:454; Rankin 2003:186; Labov 2007; etc.). However, this conclusion does not appear necessary, because a tree-like structure is nothing more than a special case of a linkage – an exceptional case in which isoglosses just happen to be nested, and temporally ordered from broadest to narrowest.

And indeed, an important strength of the wave approach is its ability to represent not only cases of crosscutting isoglosses, but also so-called “tree-like” situations when this is in fact appropriate. Imagine that, in *Figure 3* above, the members of the AB subgroup were found to share no innovation at all with the other members of the family: this is shown by the absence of any isogloss involving A, B or AB together with other languages. Such an observation may reflect the fact that the ancestors of modern speakers of A and B isolated themselves from the rest of their family, whether due to social attitudes or to physical constraints – including migration with loss of contact. What would then obtain is precisely the sort of neat social split that is represented all the time by trees.

Would such social-split signals justify preserving the Tree Model? Not necessarily, for two reasons. First, even if the existence of a separate AB cluster could be represented visually by a ‘branch’ linking Proto-ABCDEFGH to Proto-AB, the entangled isoglosses among CDEFGH would still be incompatible with a tree, and would need to be represented by waves anyway. All in all, a wave diagram such as *Figure 3* is both necessary and sufficient to display the splits in question, and a tree would add nothing more.

The second argument is of a more epistemological nature, and still favours the Wave Model even in situations of neat social split. Under the Tree Model, splits are assumed to be the only force underlying the formation of subgroups; this constitutes an aprioristic axiom for the whole model to hold together. By contrast, under a Wave approach, the identification of such splits is an empirical – and falsifiable – result of observation. In terms of historical reconstruction, this is an invaluable advantage of the latter method. In other words, Waves are not only better designed than Trees for tackling entangled situations of dialect continua and linkages; they even do better at detecting cases of neat split, which the cladistic model merely takes for granted.

### **3.5 Synthesis: Two competing models of language diversification**

In sum, trees and waves constitute two competing attempts at representing the same thing, namely historical events of early language-internal ‘horizontal’ diffusion, apprehended through the traces they left in modern languages, via later ‘vertical’ transmission. Both approaches are equally concerned with diffusion (*shared innovations*) and with transmission (*shared inheritance*). They target the very same domain (pace Campbell & Poser 2008:399), and it is indeed *genealogical* relations that I claim are better represented by waves than by trees.

Insofar as the Wave Model is agnostic as to whether genealogical subgroups should be expected to be nested or to intersect, it constitutes a more encompassing and flexible view of language diversification than the Tree Model; the latter approach entails a number of assumptions and simplifications which are not warranted by what we now know of the actual life of languages. In lieu of trees, historical linguists should use the Wave Model – or some approach derived from it – to achieve a more exact and realistic representation of the genealogical structure of the world’s language families.

## **4 A MODEL FOR CAPTURING INTERSECTING SUBGROUPS**

What we need then is a method for identifying and representing genealogical subgroups when they intersect. Among several existing proposals for non-cladistic models (§4.1), this final section will focus on one possible way of formalising the Wave Model: Historical Glottometry.

### **4.1 Alternative approaches to genealogy**

One possible reason why trees have remained pervasive in historical linguistics, despite their long-recognised flaws, is a relatively trivial one: namely, that they offer a visually elegant and easy-to-read graphical representation of a simple scenario. For the more realistic wave-based approach to ever be fully rehabilitated, then, it is necessary to design a model that readily lends itself to readability and straightforward interpretation, without compromising empirical accuracy.

Various proposals have been made to address the flaws inherent in Schleicher's *Stammbaum*. In recent years, several phylogenetic studies have tackled the issue of weakly defined subgroups, by using Bayesian maximum-likelihood methods to assess the degree of support for each subgroup in a tree (e.g. Dunn *et al.* 2008; Greenhill & Gray 2009; Greenhill, Drummond & Gray 2010; Gray, Bryant & Greenhill 2010; Bowern & Atkinson 2012; see Dunn, this volume). These welcome methods avoid a simplistic reading of family trees, and provide empirical ways to gauge the validity of tree-based genealogical hypotheses. Yet these are still cladistic approaches: faced with a linkage-type family, they can quantify the degree to which the family is “(non-)tree-like”; but they do not provide a convincing alternative representation of their own, which could be used to identify the precise patterns of intersection between genealogical subgroups.<sup>11</sup>

Other proposals have been more clearly inspired by wave- or network-based representations: Southworth's (1964) “tree-envelopes”; Anttila's (1989:305) isogloss map of major Indo-European subgroups; Hock's (1991:455) “truncated octopus-like tree”; van Driem's (2001:403) “fallen leaves”; Forster, Toth & Bandelt's (1998:185) “evolutionary network”; Ross' (1997:223, 234) social-network representations of language change, etc. Each proposal contributes to the search for a representation of language genealogies that is free from the constraining assumptions of the Tree Model. However, most of them are intuitive and programmatic, and have not been applied to detailed empirical data from specific language families.

An exception must be made for the network representations in Forster *et al.* (1998) mentioned above, as well as for NeighborNets, which have recently gained wide acceptance (Bryant, Filimon & Gray 2005; Heggarty *et al.* 2010). These networks are capable of displaying pairwise distances among taxa in the form of intersecting groupings. Free from the assumptions of trees, NeighborNets make it possible to visually capture the tangled webs typical of most language families, particularly linkages. An example of such a NeighborNet will be presented, and criticised, in §4.3.4.

Among various other approaches to representing language diversity, one should also mention *dialectometry* (Séguy 1973; Guarisma & Möhlig 1986; Goebl 2006; Nerbonne 2010; Szemrecsanyi 2011). This family of methods is used to visualise pairwise linguistic distances across a dialect continuum, calculated on the basis of large amounts of data; its results typically take the form of choropleth maps. Inspiring though it is, this approach does not attempt to tackle language history *per se*: following accepted practice among dialectologists, its assessment of linguistic distance is based merely on the comparison of synchronic features, without distinguishing shared inheritance from *shared innovations*.

## 4.2 Crossing the Comparative Method with the Wave Model: Historical Glottometry

In the final part of this chapter, I propose a synthesis of the theoretical principles discussed earlier, and outline a new model I call *HISTORICAL GLOTTOMETRY*. This method aims at combining the precision and realism of dialectological approaches (especially dialecto-

metry, from which its name is inspired) with the reasoning power of the Comparative Method. The objective of Historical Glottometry is to identify genealogical subgroups in a language family, and measure their relative strengths so as to assess their historical patterns of distribution across social networks. Stronger linguistic ties can then be taken as indicators of stronger bonds among past societies – precisely the sort of invaluable insight language historians hope to achieve.

Because the model here defined is meant to capture the unfolding of historical events which underlie language diversification, the focus of our attention needs to be not just on the synchronic properties of languages, but on those properties that are thought to reflect *shared innovations* – in accordance with Leskien’s principle (see §2.1). This key principle of the Comparative Method can perfectly well be applied to a wave-based or network-based approach: this is how, for example, *Figure 3* above should be interpreted, with each isogloss corresponding to one or more shared innovations.<sup>12</sup>

The tools for distinguishing innovations from retentions are also those of the Comparative Method, and will be illustrated in §4.3.2 below; they include the principle of regularity in sound change, hypotheses on the direction of change and on relative chronology, among other principles. In this respect, the procedure is identical to the one used to identify innovations in a cladistic approach. Likewise, the Comparative Method has often proven capable of distinguishing, for example based on the observation of regular and irregular sound changes, which properties were inherited or acquired early in a dialect continuum, and which ones were acquired later by contact across already separated languages (e.g. Biggs 1965 for Rotuman). Such tools are powerful for isolating the relevant genealogical data that will feed into our historical argumentation.

Once a number of innovations have been identified, one can observe which languages have evolved together over time. Whenever a group of languages share together one or several innovations that can reasonably be assigned to processes of language-internal diffusion, they define a (more or less well-supported) genealogical subgroup (§3.2). For each subgroup, its number  $\varepsilon$  of “exclusively shared innovations” is a measure of how frequently its members tended to imitate each other’s speech (as opposed to that of non-members), and provides a first approximation to the strength of their social bonds. For example, in *Figure 3* above, suppose that languages E-F shared 32 innovations, and F-G just 8: such a linguistic measure would show that the community F had much stronger social bonds with E than it had with G.

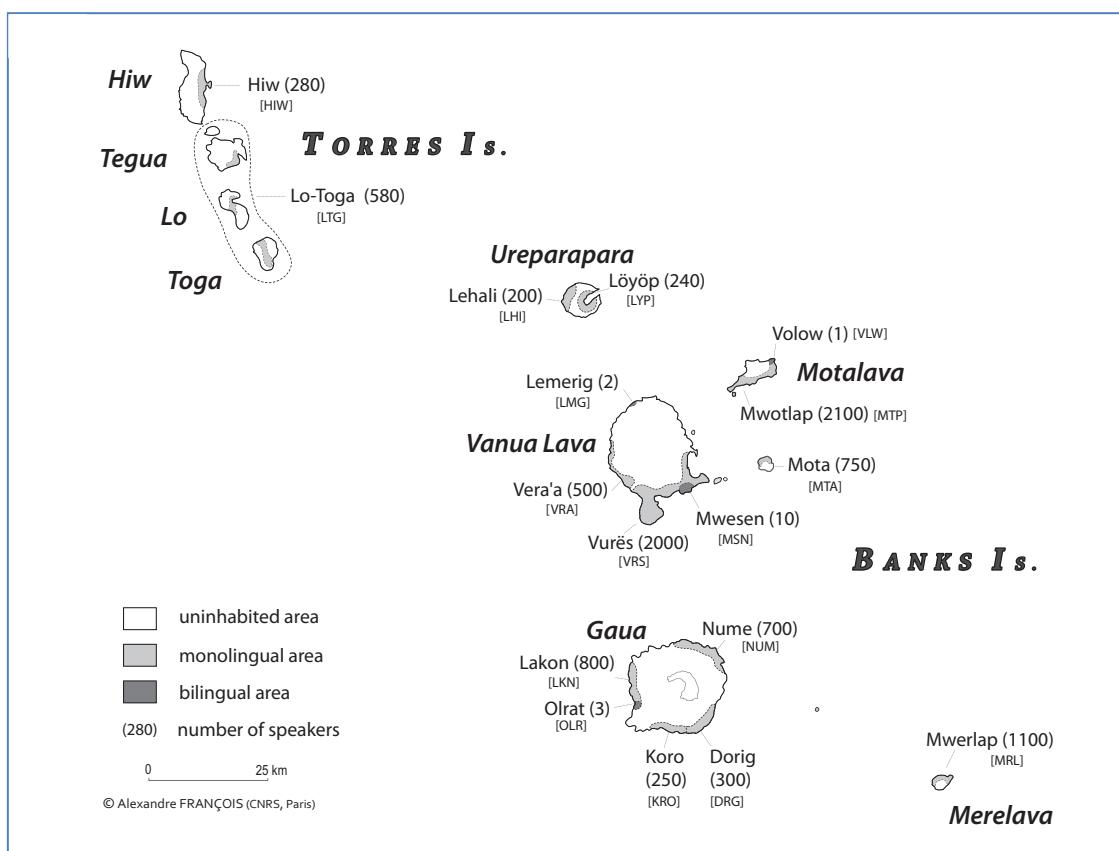
Historical Glottometry (as described in greater detail in Kalyan & François, f/c) provides still more precise tools to measure the relative strengths of subgroups in a linkage situation – in particular, calculations of *cohesiveness* and *subgroupiness*. These will be briefly presented below, based on actual data taken from the languages of northern Vanuatu.

## 4.3 A glottometric study of the northern Vanuatu linkage

### 4.3.1 THE NORTHERN VANUATU LINKAGE

Vanuatu, an archipelago of island Melanesia in the South Pacific, is home to 113 indigenous languages. They all descend from Proto Oceanic (POc), a language that was spoken about 3,200 years ago by those who first settled most of the islands of the Pacific (Pawley 1999). Apart from three Polynesian languages which arrived in Vanuatu in the last millennium, the remaining 110 languages form a linkage (Tryon 1996, Lynch 2000:181, François 2011b): their modern diversity results from three millennia of *in situ* fragmentation, with no notable external input. This diversification was brought about by the accumulation of partially overlapping isoglosses among what started as a vast dialect network, and progressively became the linguistic mosaic we know today.

Among these 110 languages, 17 are spoken in the Torres and Banks Islands in the north of the country, by a population which has always sustained traditions of interisland marriage and social contact of various kinds (François 2011a, 2012). The names of these 17 languages are given on *Map 1*, together with customised abbreviations and numbers of speakers.



Map 1: The 17 languages of the Torres and Banks Islands, in northern Vanuatu

### 4.3.2 APPLYING THE COMPARATIVE METHOD

Based on primary data I have been collecting on these 17 languages since 1997, I identified regular sound correspondences among them, and created a database of morphological and lexical reconstructions (François 2005, 2013).

The steps involved in applying the Comparative Method should be familiar to historical linguists, since most are also practised with more classical (tree-based) approaches to subgrouping. Data collected in modern languages are analysed in light of regular sound correspondences, so as to identify cognate sets and reconstruct corresponding proto-forms. For each property considered in a given language, it is possible to make reasonably solid hypotheses about whether that property is conservative of earlier stages such as Proto Oceanic, or results from a local innovation that took place – that is, emerged and diffused – after the initial settlement of Vanuatu.

For example, consider the modern forms for the verb ‘steal’ in the Torres–Banks languages (ranked geographically from northwest to southeast):

- (1) ‘steal’: *HIW βeney*; *LTG βəney*; *LHI pvl*; *LYP pil*; *VLW <sup>m</sup>bel*; *MTP <sup>m</sup>bel*; *LMG pœl*; *VRA <sup>m</sup>bɔl*; *VRS <sup>m</sup>bœl*; *MSN pɔl*; *MTA pal*; *NUM <sup>m</sup>bal*; *DRG <sup>m</sup>baɪl*; *KRO <sup>m</sup>beāl*; *OLR pal*; *LKN pal*; *MRL <sup>m</sup>bɔl*.

Knowledge of historical phonology in this region makes it clear that the two Torres forms (*HIW βeney*; *LTG βəney*) are regular reflexes of *\*panako* ‘steal’, the etymon reconstructed at the level of Proto Oceanic (Blust 2013). While these two languages exhibit sound change here, they are lexically conservative: these forms thus constitute, for the present purpose, a case of shared retention, and should not count towards subgrouping. By contrast, the forms in the 15 Banks languages all reflect an etymon which can be reconstructed, based on regular sound correspondences, as *\*<sup>m</sup>balu* (François 2005:493). This form is unattested elsewhere in Oceanic, and can therefore safely be flagged as a local lexical innovation shared by the 15 Banks languages. Doing so does not necessarily require positing a unitary “proto-Banks” language sharply divided from the rest, like a node in a tree: rather, what is defined here is simply a cluster of 15 communalects which, at some point prior to becoming completely mutually unintelligible, happened to share certain innovations together. (In fact, that cluster is sometimes crosscut by certain isoglosses: see *Table 1* below.) The identification of innovations requires that variants can be ordered in time: in this case, there is enough evidence to show that *\*panako* predates *\*<sup>m</sup>balu*, so the latter is innovative. This procedure sometimes involves reasonings on the relative chronology of changes, whenever this is justified by the data (see François 2011a:200).

Once each historical innovation was identified following similar procedures, it became possible to indicate which languages reflect it, and which don’t. At this point, I deliberately avoided making judgments – which would have been largely arbitrary – regarding whether a given innovation was a “common” or an “uncommon” type of change. While this precaution is made necessary by an all-or-nothing approach such as the Tree Model (where an uncommon change can serve as a fatal counterexample to a particular subgrouping hypothesis), it is much less relevant in a model capable of handling

innovations in conflicting distributions. In fact, in the event that a subgroup AB were supported by ten ‘rare’ innovations and BC by ten ‘common’ ones, there would be no legitimate reason for considering AB to be more strongly supported than BC: the two subgroups should be given equal weight, regardless of the nature (common vs. uncommon) of their internal innovations.

Likewise, I made no attempt to separate shared innovations from changes that potentially could have been innovated independently in two languages (*parallel innovations*), because this too could only be open to speculation. My hypothesis, which proved successful, was that a large enough number of data points should yield a strong genealogical signal based on well supported subgroups, whereas any noise due to parallel innovations would be reduced, due to the low attestation of associated language clusters.

In sum, the key to meaningful results was to first create a large database of historical innovations.

#### 4.3.3 COMPILING A DATABASE OF INNOVATIONS

I compiled a database of 474 different innovations. These include 21 instances of regular (i.e. systemic) sound change, 116 of irregular (i.e. lexically-specific) sound change, 91 of morphological change, 10 of syntactic change and 236 of lexical replacement.

For each language L and each innovation *i*, the database has ‘1’ when language L reflects *i*; ‘0’ when there is positive evidence that L did not undergo *i*; and a blank whenever the evidence is inconclusive either way. Altogether, the database contains 8058 data points: 2728 positive (‘1’), 5040 negative (‘0’) and 290 agnostic (‘–’).

*Table 1* displays a small sample of nine such innovations taken from the database, and shows their distribution across the 17 members of the linkage. Each innovation is here identified using a number (first column) and a label (second column) used here simply as a mnemonic for housekeeping purposes.

id	HIW	LTG	LHI	LYP	MTP	VLW	LMG	VRA	VRS	MSN	MTA	NUM	DRG	KRO	OLR	LKN	MRL
1	* <sup>m</sup> balu	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	*late	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0
3	*suRi	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
4	*o <sup>ŋ</sup> ga	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0
5	*ira	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0
6	*t>?	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
7	*one	0	0	0	0	0	0	1	1	1	1	–	1	1	1	0	0
8	*wo	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0
9	*goRo	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

*Table 1 – Intersecting isoglosses among Torres and Banks languages: a small sample*

The nine innovations of *Table 1* are defined as follows:

1. [\*<sup>m</sup>balu] LEXICAL REPLACEMENT  
POc \*panako ‘steal’ was replaced with a new verb \*<sup>m</sup>balu (see above)

2. [\*late] LEXICALLY-SPECIFIC SOUND CHANGE  
\*late ‘break s.th. in two’ irregularly changed to \*lete  
(e.g. VRA *lī?* is a regular reflex of \*lete but not of \*late)
  3. [\*suRi] MORPHOLOGY  
POc verb \*suRi ‘follow’ grammaticalised into a Dative preposition  
(e.g. MTP *hij*, a regular reflex of \*suRi, encodes Dative: François 2001:683)
  4. [\*oŋga] LEXICALLY-SPECIFIC SOUND CHANGE  
POc \*waŋga ‘canoe’ irregularly changed to \*oŋga  
(e.g. MTP *ɔk* is a regular reflex of \*oŋga but not of \*waŋga)
  5. [\*ira] LEXICALLY-SPECIFIC SOUND CHANGE  
\*ura ‘lobster’ (<POc \*qurāŋ) irregularly changed to \*ira (François 2011a:200)  
(e.g. LYP *n-iɛj* is a regular reflex of \*ira but not of \*ura)
  6. [\*t>?] REGULAR SOUND CHANGE  
\*t regularly changed to glottal stop \*?
  7. [\*one] LEXICALLY-SPECIFIC SOUND CHANGE  
\*eno ‘lie down’ (<POc \*qenop) metathesised to \*one  
(e.g. LMG *œn* is a regular reflex of \*one but not of \*eno)
- Note:* The etymon \*qenop has been lost altogether in Mota, where ‘lie down’ is a non-cognate form *rsa*. This lexical replacement makes it impossible to empirically assess whether pre-Mota had earlier kept the conservative form \*eno (coded as ‘0’) or undergone the metathesis to \*one like its neighbours (coded as ‘1’). Therefore I choose to remain agnostic and mark this language as one where the presence of the innovation cannot be assessed at all (coded as ‘-’). Historical Glottometry assigns a separate status to such data points, and treats them differently from 0 or 1.
8. [\*wo] MORPHOLOGY  
innovative clitic \*wo replaced the NP article \*na for alienable non-human nouns (François 2007)
  9. [\*ŋoRo] LEXICAL REPLACEMENT  
POc \*matiruR ‘sleep’ was replaced by \*ŋoRo, etymologically ‘snore’.

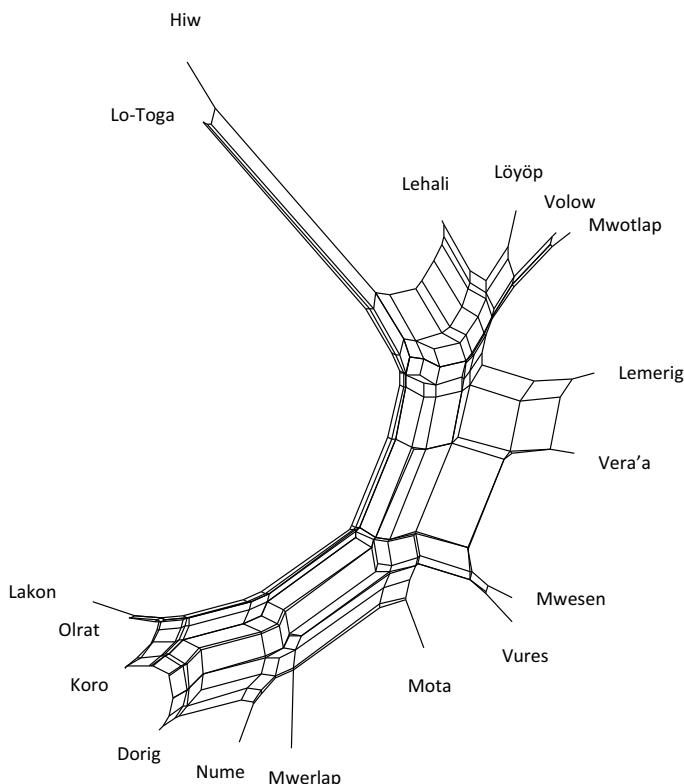
Importantly, all the innovations considered here are unlikely to result from recent borrowing, and can be safely assumed to have been diffused in the earlier times of mutual intelligibility: they are therefore strongly diagnostic of genealogical relations in the sense of the Comparative Method. This is true of cases of lexical replacement selected here, because they involve basic vocabulary items, and because the replacement evidently predated regular sound change in each language (e.g. Lakon has \*ŋoRo > ŋɔ: ‘sleep’, with regular loss of \*R and compensatory lengthening, cf. François 2011b:150). Instances of *lexically-specific sound change* are also strongly indicative of genealogy, because they are unlikely to diffuse across separate languages: these arbitrary alterations of word forms normally diffuse only across individuals who self-identify as speakers of the same language at the time of the change (Ross 1988:12; François 2011a:200).

As the table suggests, plotting innovations on a map of Torres–Banks languages typically yields patterns of intersecting isoglosses, along the lines of *Figure 3* above. Their

linguistic history cannot be represented by a tree, and is better approached using a non-cladistic model.

#### 4.3.4 DISPLAYING RESULTS IN A NEIGHBORNET

My collaborator Siva Kalyan and I used the database described above as the basis for a number of calculations, in order to identify genealogical subgroups and assess their relative strengths. *Figure 4* provides a NeighborNet of northern Vanuatu languages, based on rates of pairwise “ACQUIRED SIMILARITY” or “cohesiveness” (number of innovations shared between two languages, as a proportion of the total number of innovations affecting either one).



*Figure 4 – A NeighborNet diagram of northern Vanuatu languages, based on rates of acquired similarity*

Because the input data underlying this figure was carefully selected as representing historical *innovations* – not retentions – the distance separating any two languages reflects the accumulation of innovations over time, on one or the other side of a split. In many cases, the language clusters thus defined correspond to genealogical subgroups, as defined by a number of shared innovations. For example, *Figure 4* reflects the fact that Lemerig belongs simultaneously to two intersecting subgroups: (1) the subgroup *Lehali–Löyöp–Volow–Mwotlap–Lemerig*; and (2) the subgroup *Lemerig–Vera'a–Mwesen–Vurës*.<sup>13</sup> In that sense, NeighborNet offers promising potential for what we are looking for, namely a model for handling and representing intersecting genealogical subgroups.

However, NeighborNet has the disadvantage of being ambiguous as to which of the two sides of a split (bundle of parallel lines) corresponds to a genealogical, innovation-defined

subgroup. For example, the major split visible between Mota and Mwerlap is indicative of a genealogical subgroup, but doesn't specify which side is innovative:<sup>14</sup> one needs to look up the historical data separately to realise that the relevant subgroup here is the southern one, running from Mwerlap to Lakon.<sup>15</sup> As for the long branch at the top of the figure, it turns out that it encodes one genealogical subgroup on either side: the two Torres languages on the northwestern side (defined by  $\varepsilon=15$  exclusively shared innovations), and the fifteen Banks languages to the southeast (with  $\varepsilon=13$ ); however, this symmetrical structure is not made explicit in the figure.

Furthermore, some of the most prominent splits in this network are actually illusory, because neither side corresponds to any innovation-defined subgroup. For example, the split that runs between Lemerig and Vera'a does not correspond to any isogloss that would encompass either the languages on the northern side (Lemerig to Lo-Toga/Hiw) or those on the southern side (Vera'a to Lakon). In spite of the advantages of NeighborNet, this is, in my view, a major problem if we want to represent genealogical relations in a way that is faithful to the results of the Comparative Method.

#### 4.3.5 THE GLOTTOMETRIC ANALYSIS

The approach developed by Kalyan & François (f/c) as Historical Glottometry operates not just on pairs of languages, but on clusters of any size. This is a characteristic it shares with the Tree Model, which also deals with subgroups of various sizes; the only difference is that Historical Glottometry is capable of handling genealogical subgroups even when they intersect.

A genealogical subgroup is defined (§3.2) as any cluster of languages which have undergone at least one innovation together, at a time when they were still mutually intelligible. In this respect, any historical isogloss potentially defines a subgroup. However, defining subgroups based on weak evidence may run the risk, in some cases, of counting parallel innovations or other accidental resemblances. To avoid this pitfall, Historical Glottometry proposes a method for weighing the amount of evidence supporting each subgroup, so as to reconstruct the most significant patterns in the genealogical history of a language family.

##### 4.3.5.1 Cohesiveness

The absolute number  $\varepsilon$  of exclusively shared innovations is not the only useful measure of a subgroup's strength. Another way to assess it is to calculate the subgroup's *cohesiveness* (Kalyan & François, f/c). This measure (which I have also referred to as "acquired similarity") is relative rather than absolute: it represents the *proportion* of evidence supporting that subgroup with respect to the entire set of relevant evidence.

For any given subgroup  $G$ , let  $p$  be the number of supporting innovations (i.e. innovations which include that whole subgroup in their scope, whether exclusively or not), and  $q$  the number of conflicting innovations (i.e. innovations whose scope crosscuts  $G$ , by involving only some members of  $G$  together with some non-members). The total amount

of evidence that is relevant for assessing the cohesiveness of  $G$  is  $(p + q)$ . Now, if we call  $k_G$  the cohesiveness value of  $G$ , we have:

$$k_G = \frac{\text{number of supporting innovations}}{\text{total number of relevant innovations}} = \frac{p}{(p + q)}$$

Given any cluster of languages, cohesiveness is a measure of how close it is to a perfectly cohesive subgroup. In an ideal tree such as *Figure 2* above, subgroups are never contradicted by intersecting innovations, and their cohesiveness rate is necessarily 100%. But this rate is hardly ever met with in real-life linkages, where innovations commonly intersect.

The two languages Lemerig and Vera'a share 134 innovations – including  $\varepsilon = 9$  which they share exclusively (cf. #6 in *Table 1*). Conversely, 30 innovations are shared by Lemerig with languages other than Vera'a (cf. #4 in *Table 1*); and 14 are shared by Vera'a with languages other than Lemerig. In other words, the cohesion of the language pair *Lemerig–Vera'a* is confirmed  $p = 134$  times, but betrayed, as it were,  $q = 44$  times. The cohesiveness rate of this subgroup is thus  $k_{LMG–VRA} = 134/(134 + 44) = 0.75$ : this means that, whenever either of its members shared an innovation with at least one other language, then, 75% of the time, the isogloss encompassed both languages, thus confirming this particular subgroup. This figure can be compared with the cohesiveness of the pair *Vera'a–Vurës*, on the same island, which forms a subgroup defined by a single exclusively shared innovation. For this subgroup,  $p = 118$  and  $q = 88$ ; so  $k_{VRA–VRS} = 118/(118 + 88) = 0.57$ : that is, among the many isoglosses that affected either of the languages in this pair, only 57% involved its two members together.

From this comparison, we can make the inference that the ancestors of the Vera'a community had stronger linguistic ties – and by extension, social bonds – with Lemerig to their north ( $k_{LMG–VRA} = 75\%$ ), than with Vurës to their south ( $k_{VRA–VRS} = 57\%$ ) – in spite of the close social ties between today's Vera'a and Vurës communities. Such a metric provides a unique window onto the social networks of the past, based on the traces they left upon modern languages.

#### 4.3.5.2 Subgroupiness

In sum, the degree of support for a genealogical subgroup can be measured in two ways. In absolute terms, its number of *exclusively shared innovations* ( $\varepsilon$ ) indicates the number of times the subgroup is ‘attested’; in relative terms, its *cohesiveness rate* ( $k$ ) indicates how close it is to a perfect subgroup.

These two figures, which constitute equally legitimate measurements of a subgroup's supportedness, are mutually independent. A subgroup for which both  $k$  and  $\varepsilon$  are high is obviously strongly supported: this is the case with the pair *Mwotlap–Volow*, for example, for which  $k = 92\%$  and  $\varepsilon = 14$ . By contrast, the subgroup *Vurës–Mwesen–Mota–Nume–Mwerlap* has both low cohesiveness ( $k = 29\%$ ) and low attestation ( $\varepsilon = 2$ ): it thus has comparatively low support. But certain subgroups are only low on one of these dimensions, and therefore qualify for an intermediate level of support. For example, the pair of

languages *Dorig-Koro* has high cohesiveness ( $k = 78\%$ ), but is only attested, in my current database,  $\varepsilon = 5$  times. Symmetrically, the whole Banks subgroup – encompassing all 15 languages from Lehali to Lakon – has low cohesiveness ( $k = 30\%$ ), yet is confirmed by many isoglosses ( $\varepsilon = 13$ ).

Ideally, there would be a way to take into account not just one of these two measures, but both of them, as part of an overall assessment of a subgroup's level of support. And indeed, Historical Glottometry proposes to combine  $\varepsilon$  and  $k$  into a single figure: the absolute number of exclusively shared innovations, weighted by the subgroup's cohesiveness. This new metric, called *subgroupiness* (sigma  $\varsigma = \varepsilon \times k$ ), indicates the overall strength of the support for a given subgroup (Kalyan & François f/c).

*Table 2* displays subgroupiness values for those northern Vanuatu subgroups which have been mentioned in this chapter.

*Table 2: Measures of cohesiveness ( $k$ ) and subgroupiness ( $\varsigma$ ) of a few Torres–Banks subgroups*

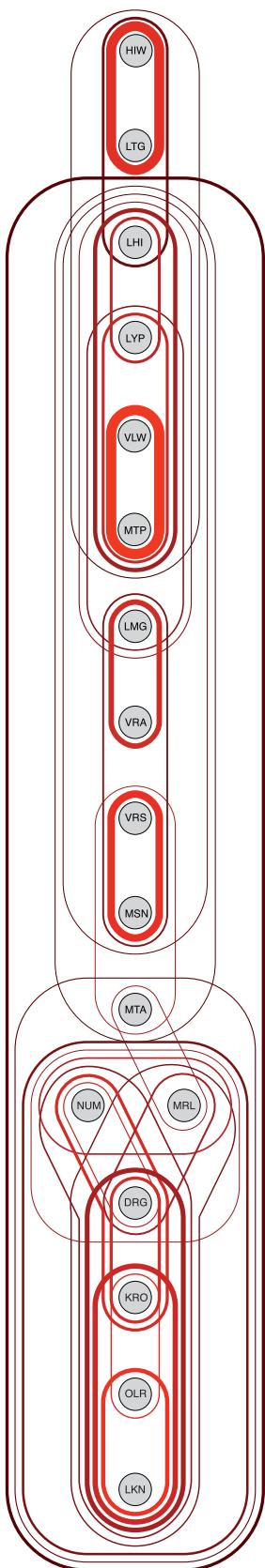
subgroup	$\varepsilon$	$k$	subgroupiness ( $\varsigma$ )
MTP-VLW	14	0.92	$\varsigma = 14 \times 0.92 = 12.82$
Hiw-LTG	15	0.83	$\varsigma = 15 \times 0.83 = 12.45$
LMG-VRA	9	0.75	$\varsigma = 9 \times 0.75 = 6.75$
DRG-KRO	5	0.78	$\varsigma = 5 \times 0.78 = 3.90$
whole Banks subgroup	13	0.30	$\varsigma = 13 \times 0.30 = 3.90$
MRL-NUM-DRG-KRO-OLR-LKN	7	0.43	$\varsigma = 7 \times 0.43 = 3.00$
LMG-VRA-VRS-MSN	5	0.44	$\varsigma = 5 \times 0.44 = 2.20$
LHI-LYP-VLW-MTP-LMG	3	0.42	$\varsigma = 3 \times 0.42 = 1.26$
VRA-VRS	1	0.57	$\varsigma = 1 \times 0.57 = 0.57$
VRS-MSN-MTA-NUM-MRL	2	0.29	$\varsigma = 2 \times 0.29 = 0.58$

#### 4.3.5.3 A glottometric diagram

Kalyan & François (f/c) calculated subgroupiness rates for all 142 attested subgroups of the Torres–Banks area. Among these, the 32 best supported ones (i.e. those above an arbitrary threshold of  $\varsigma \geq 1$ ), were brought together into a single figure, named a **glottometric diagram** (*Figure 5*). The support for each subgroup is visually represented by having line thickness proportional to *subgroupiness* ( $\varsigma$ ). The brightness of the contour line is proportional to *cohesiveness* ( $k$ ), with more cohesive subgroups appearing brighter.

This result would warrant more commentary than is possible in this paper; I will stick to the essential aspects. First of all, the subgroupiness values, as well as the diagram derived from them, confirm the statement in §4.3.1, that northern Vanuatu languages form a *linkage* in which isoglosses, and hence subgroups, constantly intersect. For example, in line with the NeighborNet above, Lemerig [LMG] subgroups both with the four languages to its north ( $\varsigma = 1.26$ ) and with the three languages to its south ( $\varsigma = 2.20$ ). Similarly Mota (MTA) forms the bridge, as it were, between a northern Banks subgroup (running from Lehali to Mota,  $\varsigma = 1.03$ ) and a distinct southern Banks subgroup (running from Mota to Lakon,

Figure 5: A glottometric diagram of the Torres–Banks languages



$\zeta = 1.30$ ). The whole island of Gaua, finally, constitutes the epitome of a perfect dialect chain.

It is worthy of notice that the glottometric approach can also detect and represent those situations which are “tree-like” (see §3.4): for example, Volow and Mwotlap clearly form a subgroup apart from Löyöp; Hiw and Lo-Toga also belong together. Yet evidently, these tree-like patches are a rarity in a language network which is strongly non-tree-like.

While the chaining of languages is essentially coherent with their spatial distribution, a finer grain of observation reveals certain non-trivial patterns that do more than just index geography. For example, even though Volow’s location is closer to Mota than to Löyöp (*Map 1*), the position of the three languages in the diagram shows that Volow and Mota are genealogically quite remote ( $k = 36\%$ ). Evidently, the ancient society of Mota had very few direct social interactions with its neighbour from Motalava island, and much more with the other islands located to its west – Vanua Lava – or to its south – Gaua, and even the remote Merelava with which Mota forms a genealogical subgroup, in spite of geographic distances. Such results illustrate the potential of the glottometric method for reconstructing the shape of past social networks.

Glottometric diagrams offer an alternative to the family tree for representing genealogical relations among languages. The analysis of innovations which underlies Historical Glottometry is entirely faithful to the Comparative Method; yet it relies on the Wave Model for one crucial insight, namely that genealogical subgroups may perfectly well crosscut each other. This empirical observation simply reflects the fact that a given community may develop social bonds with several other groups simultaneously.

## 5 CONCLUSION

Contrary to widespread belief, there is no reason to think that language diversification typically follows a tree-like pattern, consisting of a nested series of neat splits with loss of contact. Except for the odd case of language isolation or swift migration and dispersal, the normal situation is for language change to involve multiple events of diffusion across mutually intelligible idiolects in a network, typically distributed into conflicting iso-glosses. Insofar as these events of language-internal diffusion are later reflected in descendant languages, the sort of language

family they define – a “linkage” – is one in which genealogical relations cannot be represented by a tree, but only by a diagram in which subgroups intersect.

This form of language diversification – probably the most common in the world – requires an approach ultimately inspired by Schmidt's *Wellentheorie* and its overlapping waves. Among various such approaches which have been proposed, Historical Glottometry aims at detecting the genealogical structure of language families in a fine-grained, reliable and testable manner, by combining the strengths of the Comparative Method with a diffusionist, non-cladistic model of language diversification.

## 6 REFERENCES

- Aikhenvald, Alexandra, & R.M.W. Dixon, eds. 2001. *Areal diffusion and genetic inheritance: problems in comparative linguistics*. Linguistics. Oxford: Oxford University Press.
- Anttila, Raimo. 1972. *An introduction to historical and comparative linguistics*. Current issues in linguistic theory. New York: Macmillan.
- Biggs, Bruce. 1965. Direct and indirect inheritance in Rotuman. *Lingua* 14:383-415.
- Bloomfield, Leonard. 1933. *Language*. New York: Holt.
- Blust, Robert. 2013. *Austronesian Comparative Dictionary*, web edition [<http://www.trussel2.com/ACD>, accessed 22 May 2013].
- Bossong, Georg. 2009. Divergence, convergence, contact. Challenges for the genealogical classification of languages. In *Convergence and divergence in language contact situations*, edited by K. Braunmüller & J. House. Hamburg Studies on Multilingualism, 8. Amsterdam-Philadelphia: John Benjamins. Pp.13-40.
- Bowern, Claire. 2006. Another look at Australia as a linguistic area. In *Linguistic Areas*, edited by Y. Matras, A. McMahon & N. Vincent. Basingstoke: Palgrave Macmillan. Pp.244–265.
- Bowern, Claire. 2013. Relatedness as a Factor in Language Contact. *Journal of Language Contact* 6 (2): 411–432.
- Bowern, Claire, & Quentin Atkinson. 2012. Computational Phylogenetics and the Internal Structure of Pama-Nyungan. *Language* 88 (4):817-845.
- Bryant, David, Flavia Filimon, & Russell D. Gray. 2005. Untangling our past: Languages, Trees, Splits and Networks. In *The Evolution of Cultural Diversity: Phylogenetic Approaches*, edited by R. Mace, C.J. Holden & S. Shennan. London: UCL Press. Pp.67-83.
- Campbell, Lyle. 2004. *Historical linguistics: An introduction*. Edinburgh: Edinburgh University Press.
- Campbell, Lyle, & William J. Poser. 2008. *Language classification: History and method*. Cambridge: Cambridge University Press.
- Chambers, Jack K., & Peter Trudgill. 1998. *Dialectology*. Cambridge textbooks in linguistics. Cambridge: Cambridge University Press.
- Chappell, Hilary. 2001. Language contact and areal diffusion in Sinitic languages: Problems for typology and genetic affiliation. In Aikhenvald & Dixon, 328–357.
- Croft, William. 2000. *Explaining Language Change: An Evolutionary Approach*. Longman Linguistics Library. London: Longman.
- Dixon, R.M.W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- Drinka, Bridget. 2013. Phylogenetic and areal models of Indo-European relatedness: The role of contact in reconstruction. *Journal of Language Contact* 6 (2): 379-410.
- Dunn, Michael, Stephen Levinson, Eva Lindström, Ger Reesing, & Angela Terrill. 2008. Structural phylogeny in historical linguistics: Methodological explorations applied in Island Melanesia. *Language* 84 (4):710-759.
- Durie, Mark & Malcolm Ross (eds). 1996. *The Comparative Method Reviewed: Regularity and irregularity in language change*. Oxford: Oxford University Press.
- Enfield, Nicholas. 2003. *Linguistic epidemiology: Semantics and grammar of language contact in mainland Southeast Asia*. London: Routledge-Curzon.

- Enfield, Nicholas. 2008. Transmission biases in linguistic epidemiology. *Journal of Language Contact* Thema 2:299–310.
- Ernst, Gerhard; Martin-Dietrich Gleßgen, Christian Schmitt, & Wolfgang Schweickard, eds. 2009. *Romanische Sprachgeschichte—Ein Internationales Handbuch zur Geschichte der Romanischen Sprachen/Histoire linguistique de la Romania—Manuel international d'histoire linguistique de la Romania*. Berlin: De Gruyter Mouton.
- Forster, Peter; Alfred Toth & Hans-Jürgen Bandelt. 1998. Evolutionary network analysis of word lists: Visualising the relationships between Alpine Romance languages. *Journal of Quantitative Linguistics* 5 (3):174–187.
- Fox, Anthony. 1995. *Linguistic reconstruction: An introduction to theory and method*. Oxford: Oxford University Press.
- François, Alexandre. 2001. Contraintes de structures et liberté dans l'organisation du discours: Une description du mwotlap, langue océanienne du Vanuatu. Doctoral dissertation in Linguistics. Paris: Université Paris-IV Sorbonne. 1078 pp.
- François, Alexandre. 2005. Unraveling the history of the vowels of seventeen northern Vanuatu languages. *Oceanic Linguistics* 44 (2):443–504.
- François, Alexandre. 2007. Noun articles in Torres and Banks languages: Conservation and innovation. In Jeff Siegel, John Lynch & Diana Eades (eds.), *Language Description, History and Development: Linguistic indulgence in memory of Terry Crowley*. Creole Language Library 30. New York: Benjamins. Pp.313–326.
- François, Alexandre. 2011a. Social ecology and language history in the northern Vanuatu linkage: A tale of divergence and convergence. *Journal of Historical Linguistics* 1 (2):175–246.
- François, Alexandre. 2011b. Where \*R they all? The geography and history of \*R loss in Southern Oceanic languages. *Oceanic Linguistics* 50 (1):140–197.
- François, Alexandre. 2012. The dynamics of linguistic diversity. Egalitarian multilingualism and power imbalance among northern Vanuatu languages. *International Journal of the Sociology of Language* 214: 85–110.
- François, Alexandre. 2013. Shadows of bygone lives: The histories of spiritual words in northern Vanuatu. In *Lexical and structural etymology: Beyond word histories*, edited by R. Mailhammer. Studies in Language Change. Berlin: DeGruyter Mouton. Pp.185–244.
- Garrett, Andrew. 2006. Convergence in the formation of Indo-European subgroups: Phylogeny and chronology. In *Phylogenetic methods and the prehistory of languages*, edited by P. Forster & C. Renfrew. Cambridge: McDonald Institute for Archaeological Research. Pp.139–151.
- Geraghty, Paul A. 1983. *The History of the Fijian languages*. Oceanic Linguistics Special Publication, 19. Honolulu: University of Hawaii Press.
- Giles, Howard, & Tania Ogay. 2007. Communication accommodation theory. In *Explaining communication: Contemporary theories and exemplars*, edited by B. B. Whaley & W. Samter. London: Routledge. Pp.293–310.
- Gilliéron, Jules. 1880. *Petit Atlas phonétique du Valais roman (sud du Rhône)*. Paris: Champion.
- Goebl, Hans. 2006. Recent Advances in Salzburg Dialectometry. *Literary and Linguistic Computing* 21 (4): 411–435.
- Gray, Russell D., David Bryant, & Simon J. Greenhill. 2010. On the shape and fabric of human history. *Philosophical Transactions of the Royal Society London, B* 365:3923–3933.
- Gray, Russell D., Alexei J. Drummond, & Simon J. Greenhill. 2009. Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement. *Science* 323 (5913):479–483.
- Greenhill, Simon J., & Russell D. Gray. 2009. Austronesian language phylogenies: myths and misconceptions about Bayesian computational methods. In *Austronesian historical linguistics and culture history: a festschrift for Robert Blust*, edited by A. Adelaar & A. Pawley. Pacific Linguistics, 601. Canberra: Australian National University. Pp.375–397.
- Greenhill, Simon J.; Alexei J. Drummond, & Russell D. Gray. 2010. How accurate and robust are the phylogenetic estimates of Austronesian language relationships? *PLoS One* 5 (3):e9573.
- Guarisma, Gladys, & Wilhelm J.G. Möhlig, eds. 1986. *La méthode dialectométrique appliquée aux langues africaines*. Berlin: Reimer.
- Hashimoto, Mantaro. 1992. Hakka in Wellentheorie perspective. *Journal of Chinese Linguistics* 20:1–49.

- Haspelmath, Martin. 2004. How hopeless is genealogical linguistics, and how advanced is areal linguistics? *Studies in Language*, 28 (1), 209-223.
- Haspelmath, Martin, & Uri Tadmor. 2009. *Loanwords in the World's Languages: A comparative handbook*. Berlin: Mouton de Gruyter.
- Heggarty, Paul; Warren Maguire, & April McMahon. 2010. Splits or waves? Trees or webs? How divergence measures and network analysis can unravel language histories. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:3829-3843.
- Hock, Hans Henrich. 1991. *Principles of historical linguistics*. Trends in Linguistics: Studies and Monographs, 34. Berlin: de Gruyter.
- Holton, Gary. 2011. A Geo-linguistic Approach to Understanding Relationships within the Athabaskan Family. Paper read at the international workshop *Language in Space: Geographic Perspectives on Language Diversity and Diachrony*. Boulder, Colorado.
- Huehnergard, John, & Aaron Rubin. 2011. Phyla and Waves: Models of Classification. Semitic Languages. In *Semitic Languages: An International Handbook*, edited by S. Weninger, G. Khan, M. P. Streck & J. Watson. Handbücher zur Sprach- und Kommunikationswissenschaft, 36. Berlin: de Gruyter Mouton. Pp.259-278.
- Kalyan, Siva, & Alexandre François. f/c. Freeing the Comparative Method from the Tree Model: A framework for Historical Glottometry. In *Let's talk about trees: Tackling Problems in Representing Phylogenetic Relationships among Languages*, edited by R. Kikusawa & L. Reid (Senri Ethnological Studies). Osaka: National Museum of Ethnology.
- Krauss, Michael E., & Victor Golla. 1981. Northern Athapaskan languages. In *Handbook of North American Indians*, vol. 6: *Subarctic*, edited by J. Helm. Pp.67-85.
- Krishnamurti, Bh. 1998. Regularity of sound change through lexical diffusion: A study of s > h > Ø in Gondi dialects. *Language Variation and Change* 10:193-220.
- Labov, William. 1963. The social motivation of sound change. *Word* 19:273-309.
- Labov, William. 1994. *Principles of linguistic change: Internal factors*. Oxford: Blackwell.
- Labov, William. 2001. *Principles of linguistic change: Social factors*. Oxford: Blackwell.
- Labov, William. 2007. Transmission and diffusion. *Language* 83 (2):344-387.
- Leskien, August. 1876. *Die Declination im Slawisch-Litauischen und Germanischen*. Leipzig: Hirzel.
- Lynch, John. 2000. Linguistic subgrouping in Vanuatu and New Caledonia. In *Proceedings of the Second International Conference on Oceanic Linguistics (SICOL)*, vol. 2: *Historical and descriptive studies*, edited by B. Palmer & P. A. Geraghty. Pacific Linguistics, 505. Canberra: University of the South Pacific. Pp.155-184.
- Matras, Yaron, & Peter Bakker, eds. 2003. *The Mixed Language Debate: Theoretical and Empirical Advances*. Berlin: Walter de Gruyter.
- Milroy, Lesley. 1987. *Language and social networks*. Language in Society. Oxford: Blackwell.
- Milroy, James, & Lesley Milroy. 1985. Linguistic change, social network and speaker innovation. *Journal of linguistics* 21 (2):339-384.
- Nerbonne, John. 2010. Measuring the diffusion of linguistic change. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:3821-3828.
- Nettle, Daniel. 1999. *Linguistic diversity*. Oxford: Oxford University Press.
- Page, Roderick D. M. & Edward C. Holmes. 2009. *Molecular Evolution: A phylogenetic approach*. Oxford: Blackwell.
- Pawley, Andrew. 1999. Chasing rainbows: Implications of the rapid dispersal of Austronesian languages for subgrouping and reconstruction. In *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, edited by E. Zeitoun & P. J.-K. Li. Symposium Series of the Institute of Linguistics: Academia Sinica. Pp.95-138.
- Pawley, Andrew K., & Malcolm D. Ross. 1995. The prehistory of Oceanic languages: a current view. In *The Austronesians: Historical and Comparative Perspectives*, edited by P. S. Bellwood, J. J. Fox & D. Tryon, *Comparative Austronesian Project*. Canberra: Australian National University. Pp.39-80.
- Penny, Ralph John. 2000. *Variation and change in Spanish*. Cambridge: Cambridge University Press.
- Pulgram, Ernst. 1961. The nature and use of proto-languages. *Lingua* 10: 18-37.
- Ramat, Paolo. 1998. The Germanic languages. In *The Indo-European languages*, edited by A. G. Ramat & P. Ramat. London: Routledge. Pp.380-414.

- Rankin, Robert. 2003. The Comparative Method. In *The Handbook of Historical Linguistics*, edited by B. D. Joseph & R. D. Janda. Oxford: Blackwell. Pp.183-212.
- Romaine, Suzanne. 1988. *Pidgin and Creole Languages*. London: Longman.
- Ross, Malcolm. 1988. *Proto-Oceanic and the Austronesian languages of Western Melanesia*. Pacific Linguistics. Canberra: Australian National University.
- Ross, Malcolm. 1996. Contact-induced change and the Comparative Method: cases from Papua New Guinea. In Durie & Ross (eds), 180-217.
- Ross, Malcolm. 1997. Social networks and kinds of speech-community event. In *Archaeology and language 1: Theoretical and methodological orientations*, edited by R. Blench & M. Spriggs. London: Routledge. Pp.209-261.
- Ross, Malcolm. 2001. Contact-induced change in Oceanic languages in North-West Melanesia. In Aikhenvald & Dixon, 134–166.
- Schleicher, August. 1853. Die ersten Spaltungen des indogermanischen Urvolkes. *Allgemeine Monatsschrift für Wissenschaft und Literatur* 1853:786–787.
- Schmidt, Johannes. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Hermann Böhlau.
- Schuchardt, Hugo. 1885. *Über die Lautgesetze: Gegen die Junggrammatiker*. Berlin: Oppenheim.
- Séguy, Jean. 1973. La dialectométrie dans l'Atlas linguistique de la Gascogne. *Revue de linguistique romane* 145-146:1-24.
- Siegel, Jeff. 2008. *The emergence of Pidgin and Creole languages*. Oxford: Oxford University Press.
- Southworth, Franklin C. 1964. Family-tree diagrams. *Language* 40 (4):557–565.
- Street, Richard L., & Howard Giles. 1982. Speech accommodation theory: A social cognitive approach to language and speech behavior. In *Social cognition and communication*, edited by M. E. Roloff & C. R. Berger. Beverly Hills: Sage. Pp.193-226.
- Szemerécsanyi, Benedikt. 2011. Corpus-based dialectometry: a methodological sketch. *Corpora* 6 (1): 45-76.
- Toulmin, Matthew. 2006. Reconstructing linguistic history in a dialect continuum: The Kamta, Rajbanshi, and Northern Deshi Bangla subgroup of Indo-Aryan, Australian National University, Canberra.
- Toulmin, Matthew. 2009. *From linguistic to sociolinguistic reconstruction: the Kamta historical subgroup of Indo-Aryan*. Studies in Language Change, 604. Canberra: Pacific Linguistics.
- Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Blackwell.
- Tryon, Darrell. 1996. Dialect chaining and the use of geographical space. In *Arts of Vanuatu*, edited by J. Bonnemaison, K. Huffman, C. Kaufmann & D. Tryon. Bathurst: Crawford House Press. Pp.170-181.
- van Driem, George. 2001. *Languages of the Himalayas: An ethnolinguistic handbook of the greater Himalayan region*. Leiden: Brill.
- Wenker, Georg. 1881. *Sprachatlas von Nord- und Mitteldeutschland: Text und Einleitung. Auf Grund von systematisch mit Hilfe der Volksschullehrer gesammeltem Material aus circa 30000 Orten*. Strasbourg, London: Trübner.

## NOTES

<sup>1</sup> I would like to thank Siva Kalyan and Malcolm Ross for their advice on various aspects of the present chapter. This research was presented at the 21<sup>st</sup> International Conference of Historical Linguistics (ICHL21) in Oslo, in August 2013. It forms part of the research strand “Typology and dynamics of linguistic systems” of the LabEx *Empirical Foundations of Linguistics* (funded by ANR-CGI).

<sup>2</sup> See Nettle (1999). For a case study of how these opposing processes interact in a specific region of Melanesia, see François (2012).

<sup>3</sup> I follow here the proposal by Haspelmath (2004:222) to use the term “genealogical” for what have been traditionally labelled “genetic” relations, to avoid confusion with biological genetic relations. For a discussion of what is meant by *genealogy* in historical linguistics, see §3.1.

<sup>4</sup> There is sometimes ambiguity as to whether social separation is understood as the *cause* or the *consequence* of linguistic divergence. Indeed, social or physical isolation entails that dialects will develop separately from each other; but in addition, following a sort of snowball effect, the more dialects diverge, the higher the language barrier for future communication, and thus the more the social communities may be assumed to develop independently from their neighbours, as their dialects

evolve into mutually unintelligible languages. As we will see below, the latter view is quite simplistic, and communication often continues for a long time in spite of earlier events of linguistic divergence.

<sup>5</sup> I use the term *diffusion* here in its usual sense of propagation through a social network of individuals (as in Labov 2007). This is distinct from the process of *lexical diffusion*, which describes the way certain forms of sound change propagate across the lexicon (Labov 1994:421; Krishnamurti 1998).

<sup>6</sup> Hale (this volume) expresses a similar idea in terms of individual “grammars”.

<sup>7</sup> While the two processes of diffusion – language-internal vs cross-linguistic – are fundamentally similar in the way they spread through a population, they still differ in their precise psycholinguistic mechanism, and in the nature of the linguistic features they affect (Bowern 2013): for example, “basic vocabulary” items are more likely to spread through language-internal diffusion than through contact (Haspelmath & Tadmor 2009:65–68). This sort of difference is not relevant to our main point here, which is to say that in both cases, the Tree Model is ill-designed to represent the facts of diffusion – including those that define genealogical relations.

<sup>8</sup> For empirical illustrations of this point, see for example Geraghty (1983) for Fijian communalects, Garrett (2006) for ancient Greek dialects, François (2011a:201) for northern Vanuatu.

<sup>9</sup> Societies differ on how much linguistic fragmentation they tolerate. Some more centralised societies may involve a higher degree of levelling between dialects, in such a way that a change affecting the more central or influential varieties will rapidly spread to the whole network of individuals who self-identify as speakers of that “language”. Conversely, some societies are more tolerant towards internal diversity, and exert less pressure towards dialect levelling.

<sup>10</sup> In some cases, dialect levelling may erase the earlier entangled structure of a continuum, and produce the “mirage” of discrete subgroups (Garrett 2006). For example, in *Figure 3* above, should dialects E and F be wiped out as distinct varieties, then the isoglosses would appear nested again, and the family could be rendered by a tree. However, a tree-like structure is not a necessary result of dialect levelling. Thus, if the process meant the demise of dialects B, C and H in *Figure 3* but the survival of other varieties, then the genealogical structure of the linkage descended from this continuum would still resist any cladistic approach. For example, it can be shown that Italian, Spanish and French do not properly fit into a tree, even without considering the numerous intervening dialects (Kalyan & François f/c).

<sup>11</sup> Another problem is that some of the work conducted using these methods is not based on the Comparative Method. Dunn *et al.* (2008), for example, identify their subgroups based on a matrix of typological features such as word order, rather than on linguistic reconstruction and the identification of innovations.

<sup>12</sup> Because dialectologists use the term ‘isogloss’ regardless of its historical nature, one may want to specify that the isoglosses used in Historical Glottometry are all HISTORICAL ISOGLOSSES – à la Bloomfield (1933:316) or Anttila (1985:305).

<sup>13</sup> Among other relevant diagnostic innovations, the first of these two subgroups is defined by the lexically-specific change *\*wa<sup>ŋ</sup>ga* > *\*o<sup>ŋ</sup>ga* ‘canoe’ (see *Table 1*); the second by the lexically-specific dissimilation *\*mama<sup>n</sup>ri<sup>n</sup>ri* > *\*mamayi<sup>n</sup>ri* ‘cold’.

<sup>14</sup> This limitation could be rectified by including the proto-language, Proto Oceanic, as one of the taxa displayed in the NeighborNet. In this case, whichever side of the split does *not* include the proto-language, would be the group defined by innovations. The practice of including the ancestral node as a taxon in a NeighborNet, however, does not seem to be widely followed in linguistics.

<sup>15</sup> For example, these six languages share the use of a preposition *\*ma<sup>n</sup>ge* ‘above’; or the lexically-specific loss of the phoneme \*R in *\*nañoRap* ‘yesterday’ and *\*waRisa* ‘two days from now’ (François 2011b:157).

## KEYWORDS

Tree Model – Wave Model – Comparative Method – linkage – language genealogy – genetic linguistics – subgrouping – dialectology – Vanuatu – Oceanic languages

