Alexandre François*

Lexical tectonics: Mapping structural change in patterns of lexification

https://doi.org/10.1515/zfs-2021-2041

Abstract: Whether it is based on philological data or on reconstruction, historical linguistics formulates etymological hypotheses that entail changes both in form and in meaning. Semantic change can be understood as a change in "patterns of lexification", i. e., correspondences between forms and senses. Thus a polysemous word, which once lexified senses s1–s2–s3, evolves so it later encodes s2–s3–s4. Meanings that used to be colexified are now dislexified, and vice versa. Leaning on empirical data from Romance and from Oceanic, this study outlines a general model of historical lexicology, and identifies five types of structural innovations in the lexicon: split, merger, competition, shift, and relexification.

The theoretical discussion is made easier by using a visual approach to structural change, in the form of diachronic maps. Semantic maps have already proven useful to represent synchronic patterns of lexification, outlining each language's emic categories against a grid of etic senses. The same principle can be profitably used when analysing lexification patterns in diachrony: lexical change is then viewed as the reconfiguration of sense clusters in a semantic space. Maps help us visualize the "lexical tectonics" at play as words evolve over time, gradually shifting their meaning, gaining or losing semantic territory, colliding with each other, or disappearing forever.

Keywords: lexical typology, semantic change, semantic maps, structural linguistics, historical linguistics, polysemy, colexification, relexification

1 Introduction

1.1 Comparing lexical structures: From synchrony to diachrony

Systematic lexicology, pioneered by structuralists (Hjelmslev 1943; Ullmann 1957; Coseriu 1964), was revived recently through studies in lexical typology, bringing to light various ways in which lexicons can be compared across languages (see

***Corresponding author: Alexandre François,** LATTICE, CNRS–ENS–Sorbonne nouvelle, Montrouge, France; and Australian National University, Canberra, Australia, e-mail: alexandre.francois@ens.fr

Vanhove 2008a; Urban 2012; Juvonen and Koptjevskaja-Tamm 2016; Koptjevskaja-Tamm and Liljegren 2017).

When comparing how languages cut up the semantic space, one efficient approach consists in breaking up a given lexical domain into a number of potential meanings, independent of their expression in individual languages. The "etic grid" formed by these potential senses then serves as a background against which emic categories of particular languages can be compared (Koptjevskaja-Tamm et al. 2007; Evans 2010).

Several authors have advocated semantic maps as a powerful way to visualize the results of such cross-linguistic comparison (Haspelmath 2003; François 2008; Georgakopoulos et al. 2016). Until recently, most authors have adopted a synchronic approach to lexical maps (e. g., Rakhilina and Reznikova 2016: 113); however, the field shows increasing interest towards adapting the map approach to diachronic semantics (Georgakopoulos and Polis 2018; 2021).

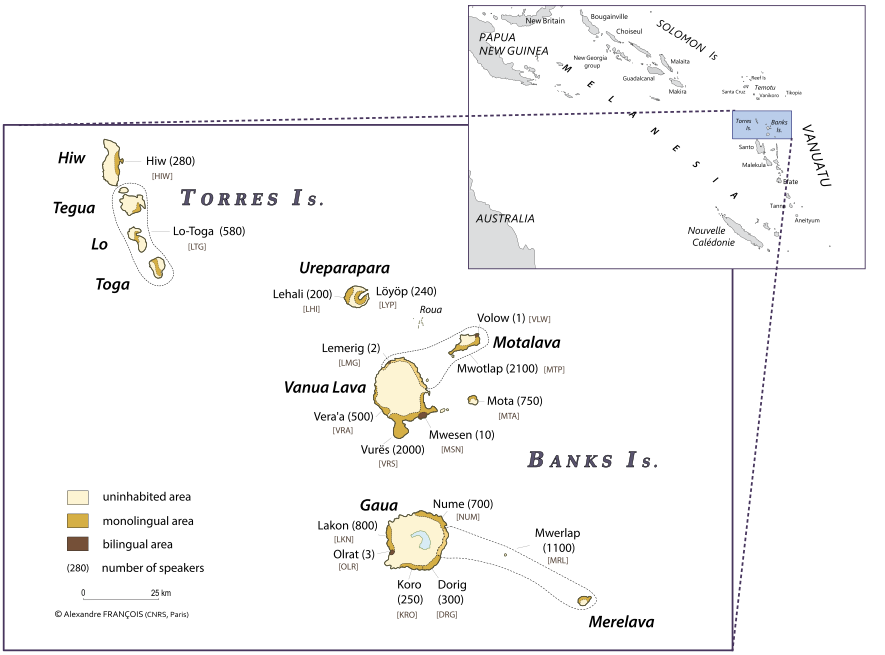
In recent decades, a number of authors have used diachronic maps as they explored the historical development of a given semantic field in particular languages – e. g., Matisoff (1978) for body parts in Tibeto-Burman; Evans (1992), Wilkins (1996) for semantic reconstruction in various Australian languages; Pawley (2005) for exploring the meanings of a Proto-Oceanic etymon; J. François (2007; 2013) for some polysemies in French and German; and so on. Coming in the wake of these authors, the present study intends to reflect on some general principles and methods pertaining to the use of maps in the study of lexical change, and to propose ways to systematize their use in historical lexicology.

As we attempt to model the evolution of lexical structures, semantic maps will appear as a powerful device to capture graphically the way in which every semantic innovation redesigns lexical domains and their internal organisation.

1.2 This study: Data sources and outline

My observations will focus on two language groups for which the history of semantic change can be solidly reconstructed: Romance and Oceanic. Romance is a family with a well-known history, and a linguistic ancestor that is well understood.

Oceanic languages, in turn, have no written tradition, and are only known in their modern forms. However, their shared ancestor Proto Oceanic (POc) can be reconstructed with reasonable confidence thanks to the Comparative method, based on its 500 modern descendants (Pawley and Ross 2006; Ross et al. 2016). Among these 500 Oceanic languages, 138 (François et al. 2015) are spoken in the Vanuatu archipelago. This study will focus on the 17 languages of the Torres and



Map 1: The languages of the Torres and Banks Islands, in northern Vanuatu.

Banks Islands of north Vanuatu (Map 1 shows their number of speakers and the three-letter abbreviations I use in this paper).

My data originates in personal fieldwork carried out on the 17 languages of the area, combining language immersion, recordings of connected speech, and linguistic elicitation by means of a personal questionnaire (François 2019).

Regular sound correspondences for northern Vanuatu languages are exposed in François (2005) for vowels, and François (2016: 31) for consonants. When a linguistic innovation is shared by Torres and Banks languages together, it can be captured by positing an intermediate common ancestor called PTB "Proto Torres–Banks" (François 2011a; 2016) – itself a descendant of Proto Oceanic. Lists of PTB lexical reconstructions appear in François (2005; 2013; 2016).

The present study will explore the diachrony of lexical structures, and propose ways to conceptualize and represent that diachrony. I will focus on *patterns of lexification* – though it must be acknowledged that other aspects, not dealt with in the present work, also contribute to a lexicon's organisation: e. g., lexical gaps, combinatorics, phraseological routines (Koptjevskaja-Tamm and Liljegren 2017; Schapper 2021).

Section 2 will first outline the methodology and conceptual framework underlying our approach. Then Section 3 will discuss instances, first of structural stability, and secondly of structural innovation – based on Romance and Oceanic data. Section 4 will discuss the phenomenon of lexical competition, a key for understanding the process of lexical replacement and structural change.

Our final discussion will propose a typology of structural innovations in the lexicon – and provide some of them with empirical illustrations from firsthand data. This will give us the opportunity to discuss a phenomenon that is frequent in northern Vanuatu: the tendency for innovative lexical structures to realign with those of their neighbours, yielding a process of *language-internal relexification*.

2 Methodology: Analyzing patterns of lexification

2.1 Patterns of lexification and semantic change

I propose to define *patterns of lexification*, taken as a synchronic property, as in (1):[1]

(1) In a given lexicon, considered at a particular point in time, *patterns of lexification* correspond to the patterns of distribution between senses and forms.[2]

In the lexicon of modern English, a word like *trunk* may refer to a portion of a tree, but also to a part of the human body, or a quite different part of an elephant; it may also refer to a large box, or a part of a vehicle; etc. The word *trunk* covers a certain semantic territory, while drawing contrasts with adjacent lexical units of the same semantic fields (e. g., *bole*, *torso*, *chest*, *snout*, *suitcase*, *boot*).

If we now take a diachronic perspective, what we call semantic change happens when those patterns of distribution between senses and forms are modified:

(2) *Semantic change* takes place when patterns of lexification (the distribution between senses and forms) undergo a change over time.

Or, as Ullmann (1957) puts it:

1 What I label "patterns of lexification" is analogous to what Coseriu (1964) called *lexical structures*.

2 In this definition, "forms" refer not only to words, but also, more generally, to conventionalised constructions, in the sense of Construction Grammar (Fillmore et al. 1988; Goldberg 2003; Barðdal et al. 2015).

> A semantic change will occur whenever a new name becomes attached to a sense and/or a new sense to a name. (Ullmann 1957: 171)

Thus the noun *corn*, which used to be a generic term for all sorts of grain, specialized historically in British English as a synonym of *wheat*, while in US English it now refers to maize. Over time, the correspondence between forms and senses has evolved.

To take a spatial metaphor, we can say that the semantic territory covered by the word *corn* has moved, as it were, across the semantic space. The present study, titled "lexical tectonics", intends precisely to build on this spatial view of semantic change, and see how it can inspire a method of data visualization that can make the most of semantic maps.

2.2 Establishing senses

Discussing semantic change amounts to analyzing how the correspondences of forms and senses evolves over time. One methodological question, then, is how we can identify *senses* in the first place.

A widespread conception is the notion that semanticists should start their work, prior to any observation, by establishing an ontology of semantic primitives, that would exist independently of their linguistic expressions. Two schools could then be followed:

- One approach, which may be called top-down, can endeavour to reach the most abstract, generic senses possible (similar to the semantic primes of Wierzbicka 1992, Goddard 2001). Then these ultra-generic semantic primitives would combine to yield specific meanings: e. g., 'boy' results from the combination of 'child' + 'male'.
- One diametrically opposed approach, this time bottom-up, could propose to break up the semantic universe into as many specific meanings as conceivable (e. g., all possible subtypes of snow, of sand, of stone...). This would compose a maximal etic grid of low-level atomistic meanings, which are grouped differently by natural languages.

Such aprioristic approaches, however, raise more questions than answers, as they depend considerably on arbitrary decisions or on the linguist's speculation. They do not appear methodologically grounded.

It is more solid to ground the definition of senses on the empirical observation of natural languages, by focusing only on those semantic contrasts that are actually attested:

(3) For the purpose of cross-linguistic comparison, a *sense* is considered *attested* if it defines the semantic contrast between two different forms, in at least one language.

The empirical observation of languages thus becomes the key to identifying a list of potential senses, which in turn can help feed their comparison.

To take an example inspired by Hjelmslev (1943), it would be speculative to wonder aprioristically how many distinct senses in total might be assigned to the domain of wood and woody materials. The Danish noun *træ* may designate a tree in the wild, or the hard matter of which it is made; it may refer to timber used in carpentry, or to the wood added to fuel a fire... If we consider that word on its own, we have no principled way of deciding whether we are dealing with a single vague meaning, or four or more "senses". But then cross-linguistic comparison shows that Danish *træ* sometimes corresponds to English *tree*, and sometimes to *wood*: this is enough to identify empirically two distinct senses (one covered by Eng. *tree*, the other one by *wood*), and just say that Danish expresses them identically. The reasoning can be recursive, as we observe that Spanish not only has a separate word for 'tree' (*árbol*), but two nouns for 'wood': *leña* 'firewood' vs. *madera* 'wood as material'. Finally, Mwotlap (Oceanic, Vanuatu) uses a single word *qētēnge* for {'tree' + 'wood as material'} but has a separate word *lēt* for 'firewood'.

In sum, we can compare the patterns of lexification across four languages by just taking into account these three attested senses, and by observing how they are distributed across lexical forms – see Table 1. Given that data set, one can account for the observed semantic diversity by acknowledging exactly these three different senses – neither more nor less.

As more lexical systems are brought into the comparison, even more patterns can be established (e. g., many languages in Australia and New Guinea use the same word for 'firewood' and 'fire' – Schapper et al. 2016), and the list of attested senses is bound to grow. But crucially, the patterns attested are not random or infinite. For every lexical domain, it is generally possible to come up with a finite list of attested senses, relevant to cross-linguistic comparison.

Table 1: Establishing a grid of etic senses based on language comparison.

Danish	English	Spanish	Mwotlap	senses
træ	*tree*	*árbol*	*qētēnge*	tree in the wild
	wood	*madera*		wood as material
		leña	*lēt*	firewood

The last column in Table 1 constitutes an *etic grid*, a list of attested senses for a given semantic field. Just like what happens in phonology, languages then differ in their *emic categories*, i. e. the language-specific lexemes that encode these senses. This is how lexical typology can incorporate a universal dimension (the etic points in the grid) while acknowledging the unique profile of each lexicon (the emic categories embodied by each lexeme).

2.3 Describing lexical structures through dis- and co-lexification

Languages may be compared in their patterns of *colexification* – that is, their ability to encode two distinct senses using the same lexeme (François 2008):

(4) A language colexifies two distinct senses if it can express them using the same lexical form.

For example, English colexifies 'wood as material' and 'firewood', because it can designate both these senses using the same word *wood*. (The notion of colexification is not whether a language must express two senses identically, but whether it can.)

Symmetrically, languages can be compared in the way they draw contrasts between close meanings. I propose to call this *dislexification*, the polar opposite of colexification:[3]

(5) A language dislexifies two senses if it must express them using different lexical forms.

Thus, Spanish dislexifies the two senses 'wood as material' (*madera*) and 'firewood' (*leña*), as it lacks any term that would encompass both meanings. As for English, it could also express that contrast, using a specific compound like *firewood*: but crucially it doesn't have to, since it can resort to the vague term *wood* in all cases (e. g. *Nights are getting cold, let's go cut some wood*). In sum, the two senses in question are dislexified in Spanish, but colexified in English.

The notion of dislexification, just like its counterpart colexification, entails no judgment on what is to be considered expected or normal. When comparing a given lexical domain across two languages, selecting the right descriptor here is essentially a matter of perspective, and is ultimately arbitrary. Likewise,

3 The contrast between prefixes *co*(*n*)- and *di*(*s*)- is reminiscent of word pairs such as *concord* vs. *discord*, *converge* vs. *diverge*, *conjoint* vs. *disjoint*.

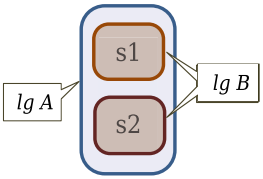


Figure 1: Two senses s1 and s2 may be colexified (*language A*) or dislexified (*language B*).

the two terms are to be understood as purely descriptive, and synchronic: co- and dis-lexification do not inherently refer to a change in meaning – contrary to "lexical merger" or "lexical split", which will be introduced later on in this study.

Together, colexification and dislexification define a structural property which can be named *patterns of lexification*;[4] see (1) above.

Lexification patterns can be compared across languages, and shown visually on a simple map such as Figure 1. A semantic domain is projected onto a two-dimensional space, in such a way that its component senses s1 and s2 occupy distinct locations on a two-dimensional plane. Using sets, senses s1 and s2 are here shown to be colexified in language A, but dislexified in language B.

While Figure 1 compares separate languages, nothing prevents us from using the same principle for comparing two historical stages of the same language. A language may dislexify two senses at a certain point in history, but its descendant may colexify them, or vice versa. This will be central to our discussions on historical lexicology.

3 Stability and change in lexical structures

Lexical structures may evolve in different directions – including the absence of change. This section will first examine cases of structural stability in Romance languages, and then turn to two cases of structural innovation in the languages of Vanuatu. These examples will be the occasion to define two important innovations in lexification patterns: lexical mergers and splits.

4 The term *clusivity* was likewise coined as a neutral term referring to the contrast between *inclusive* and *exclusive* personal pronouns (Filimonova 2005).

Table 2: 'Arm' and 'hand' in Romance languages.

	Latin	Portug.	Castilian	Catalan	French	Italian	Roman.
'arm'	*bracchium*	braço	brazo	braç	bras	braccio	braț
'hand'	*manus*	mão	mano	mà	main	mano	mână

Table 3: 'Leg' and 'foot' in Romance languages.

	Latin	Portug.	Castilian	Catalan	French	Italian	Romanian
'leg'	*crūs*	perna	pierna	cama	jambe	gamba	gambă / *picior*
'foot'	*pēs*	pé	pie	peu	pied	piede	*picior*

3.1 Structural stability

If a set of modern languages descend from a single ancestral protolanguage, part of their lexical structures will just be inherited from that ancestor. These structures can prove stable over time.

To take a simple example, we know that Latin distinguished between *bracchium* 'arm' and *manus* 'hand' – contrary to 37 percent of the world's languages, which colexify these two body parts (Brown 2013).[5] This pattern of dislexification {arm ≠ hand} has been preserved intact in modern Romance languages, all of which contrast 'arm' vs. 'hand' (Table 2). In this specific case, the inheritance of lexical structures came along with the preservation of the inherited etyma themselves: such a case can just be seen as an unremarkable instance of linguistic conservation.

Interestingly, lexical structures may remain stable even when the word itself changes its form due to lexical replacement. Consider, in Table 3, the terms for 'leg' and 'foot' in Romance languages. The word *crūs* of Classical Latin was lost in modern Romance languages, which have replaced it with new lexical material – either from Lat. *perna* 'ham', or from Late Lat. *gamba* < Gr. κάμπη 'flexure'. And yet, in doing so, they never lost the original structural pattern whereby 'leg' is dislexified from 'foot'.

5 The Database of Cross-Linguistic Colexifications "CLiCS³" (List et al. 2018) lists 300 languages colexifying 'arm' and 'hand' [https://clics.clld.org/edges/1277-1673].

The only exception here is Romanian, where *picior* 'foot' can also be used for 'leg'.[6] In modern Romanian, the inherited form *gambă* is now mostly used for the lower leg (calf, shank), while nowadays the default term for the lower limb as a whole is *picior*. Examples such as *picioare strâmbe* 'crooked legs', or *un picior mai scurt decât celălalt* 'one leg longer than the other', clearly refer to the leg rather than just the foot.

One can thus say that Romanian has innovated a colexification {leg = foot} which was not present in its ancestor. This structural innovation reflects areal influence, as Romanian calqued the colexification {leg = foot} characteristic of its neighbours: Hungarian *láb*, Bulgarian *крак*, Russian *нога*, Greek *πόδι*. Romanian is here the only exception to the general observation that Romance languages generally show structural stability in the domain of body parts for upper and lower limbs.

Many other examples could be found, in the world's languages, of inherited lexical structures being preserved over time. While this preservation sometimes goes along with the conservation of the original lexemes, we saw that it can also occur even in the case of lexical replacement – as though the structural properties of the lexicon were somehow independent of the specific words that embody them (François 2010; 2011a: 226). And indeed, our final discussion [Section 5] will examine cases of "language-internal relexification", where lexical structures remain stable in spite of lexical forms being renewed.

The structuralist Coseriu (1964) made the same point: "a semantic structure can be maintained even when the word forms have been replaced".[7] Such observations prompted him to promote what I would call *semantic lexicology* ("lexicologie du contenu", Coseriu 1964: 164) as a science of the lexicon dedicated to semantic structures per se, independent of changes in their phonological material.

3.2 Structural innovations: Mergers and splits

Languages commonly go through structural innovation. A given lexification pattern can be lost or redesigned over time. Ultimately, any semantic change in the lexicon comes with a change – whether minor or major, temporary or permanent – in the structural organisation of its semantic domain.

6 The CLiCS³ database [https://clics.clld.org/edges/1297-1301] lists 358 languages colexifying 'leg' and 'foot'.

7 In the original: "une structure sémantique peut se maintenir en dépit des remplacements des signifiants" (Coseriu 1964: 172).

Table 4: Some upper body parts in a sample of northern Vanuatu languages.

	etymon	HIW	LTG	LYP	MTP	LMG	VRS	MTA	DRG	LKN	MRL
'arm'	**lima*	mja	lime	n-pene	na-mne	pini	ᵐbeni	pane	lma	lima	lime
'hand'	**lima*	mja	lime	n-pene	na-mne	pini	ᵐbeni	pane	lma	lima	lime
'wing'	**bani*	pini	pəni	n-pene	na-mne	pini	ᵐbeni	pane	ᵐbni	pane	beni

I will illustrate the notion of structural innovation by observing two lexical domains in the Oceanic languages of northern Vanuatu: body parts and kinship terms.

3.2.1 Body part terms in northern Vanuatu

Contrary to Romance (Table 2 above), Oceanic languages generally colexify 'hand' with 'arm'. This pattern can be reconstructed for the Proto-Oceanic ancestor with a form **lima* 'arm, hand; five' (Osmond and Ross 2016: 160).

Within our northern Vanuatu sample (Table 4), all languages have preserved the colexification pattern {arm = hand}, whether they reflect the etymon **lima* (white cells) or an innovative form (coloured cells).[8] This is another case of structural conservation in spite of lexical change (Section 3.1).

Nine languages in the central area (five of which are represented in Table 4) went through lexical replacement. They lost their reflexes of **lima*, and replaced them with **bani*, whose original meaning was '[bird] wing' (Osmond and Ross 2016: 162). For the innovative languages shown in Table 4, we can reconstruct a process of semantic extension that gave rise to a new pattern of colexification:

– *stage 1*: **bani* {wing} ≠ **lima* {arm = hand}
– *stage 2*: **bani* {wing = arm = hand}

This structural change can be visualized on a semantic map (Figure 2), showing the two stages successively. Compared to Proto-Oceanic, or to the eight conservative languages that have preserved Stage 1, the nine languages that went through the lexical merger illustrated in Stage 2 have modified their lexification structures.

I propose to describe this type of change as a *lexical merger*:

8 Table 4 is ranked geographically, from Hiw in the northwest to Mwerlap in the southeast. For abbreviations of language names, see Map 1, or the Appendix. To gain space, only ten languages are cited here out of seventeen. Forms are given in IPA.

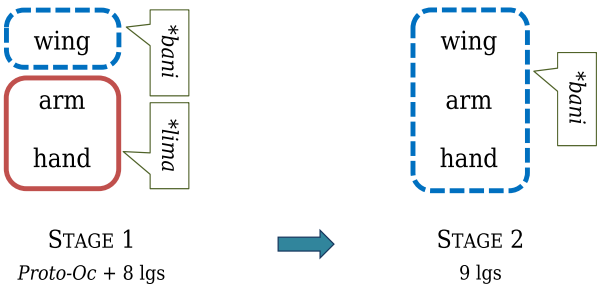


Figure 2: Terms for upper limbs in northern Vanuatu: a lexical merger.

(6) A lexical merger is a historical process of semantic change whereby two sets of senses that used to be dislexified (encoded by different lexemes) end up being colexified (encoded by the same lexeme).

Coseriu (1964) discusses similar cases of "fusion fonctionnelle", which he defines as the loss of a contrast[9] – drawing explicit analogies with structural mergers in phonology or morphology. As an example, he cites the Latin kin terms *patruus* 'father's brother' vs. *avunculus* 'mother's brother' – a contrast which modern Romance languages later merged into a single category 'uncle' (Fr. *oncle*, Ital. *zio*, Cast. *tío*...).

The opposite of a lexical merger is a *lexical split*. Table 5 lists the same sample of ten languages, this time looking at the lexification of 'leg' and 'foot'.

The general pattern in the region is the colexification {leg=foot} – a pattern itself parallel with {arm=hand} described above. Setting aside sound change, all modern forms point to the same etymon PTB **raŋo*, which can safely be reconstructed as having both meanings 'leg, foot' (François 2005: 498). But one language, Lo-Toga, stands out in the region: it is the only one that has a special word /təple/ for 'foot', distinct from 'leg' /rəŋo/. Clearly, this word results from a local innovation, which took place only in Lo-Toga and nowhere else.

Table 5: Some lower body parts in a sample of northern Vanuatu languages.

	etymon	HIW	LTG	LYP	MTP	LMG	VRS	MTA	DRG	LKN	MRL
'leg'	**raŋo*	ᵍLəŋə	rəŋo	jøŋø	na-jŋı	røŋø	røŋø	raŋo	rŋʊ	rɔŋɔ	rəŋʊ
'foot'	**raŋo*	ᵍLəŋə	təple	jøŋø	na-jŋı	røŋø	røŋø	raŋo	rŋʊ	rɔŋɔ	rəŋʊ

9 "[O]n constate la disparition d'un trait distinctif, et en conséquence la réduction de deux unités fonctionnelles à une seule unité" (Coseriu 1964: 175).

The latter type of change (structurally the opposite of a lexical merger) can be called a *lexical split*:

(7) A lexical split is a historical process of semantic change whereby two senses that used to be colexified (encoded by the same lexeme) end up being dislexified (encoded by different lexemes).

Coseriu (1964) discusses similar structural processes, and defines them as the creation of a new emic contrast ("l'apparition d'un trait distinctif nouveau", 1964: 174). He illustrates this with Lat. *avis* 'bird', a generic category which Portuguese divided in two – with *pássaro* 'small bird' vs. *ave* '(non-small) bird'.

3.2.2 Words for siblings in northern Vanuatu

The previous section defined the diachronic notions of lexical merger and lexical split, using simple examples taken from the domain of body parts in Vanuatu languages.

While remaining in Vanuatu, we can now turn to a more complex case of structural reshaping in the domain of kinship terms – namely, terms for siblings.

English divides this lexical domain based on the target's gender, contrasting *sister* vs. *brother*; but other languages use different criteria to organise the different types of siblings (Evans 2010).

A system with three categories of siblings

Let us start by observing how the field is structured synchronically in one language of the Banks islands, e. g., Lakon [LKN] from the island of Gaua. In this language, the sibling domain is organised according to two semantic features:

- the RELATIVE SEX of the target compared with that of its anchor ('ego'):
  - *tata* 'opposite-sex sibling' is the term used
    for a male's sister [♂Z]
    or for a female's brother [♀B][10]
  - *tua* and *tahi* (see below) both refer to a 'same-sex sibling',
    whether a male's brother [♂B]
    or a female's sister [♀Z]

10 For kinship abbreviations (see Dousset 2011; Hamberger et al. 2011), I use the following conventions: B 'brother'; Z 'sister'; ♂ 'male ego'; ♀ 'female ego'; |e| 'elder', |y| 'younger'. Thus ⟨♀ eB⟩ reads '[*female ego*] elder brother'; ⟨♂ yZ⟩ reads '[*male ego*] younger sister'.

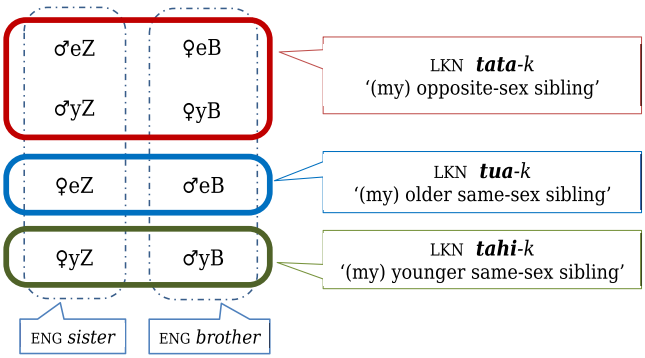


Figure 3: Terms for siblings in the Lakon language of Gaua island (*Banks group*).

- [for same-sex siblings] the RELATIVE AGE of the target compared with its anchor:
  - *tua* 'older same-sex sibling' is the term used
    for a male's elder brother [♂eB]
    or a female's elder sister [♀eZ]
  - *tahi* 'younger same-sex sibling' is the term used
    for a male's younger brother [♂yB]
    or a female's younger sister [♀yZ]

Lakon and English thus differ drastically in the way they categorize sibling relations. Their differences can be shown on a semantic map whose background is an etic grid listing a number of specific kin relations – e. g., ⟨♀yB⟩ 'a female's younger brother'. English and Lakon group these atomic senses according to their respective emic categories – see Figure 3.

The ternary structure at the source of modern Torres–Banks systems

The ternary pattern of Lakon is also found in several other languages of the Torres–Banks area, and even in several Oceanic languages spoken outside that area. This semantic organisation can in fact be safely reconstructed for PTB ('Proto-Torres–Banks'), the region's shared linguistic ancestor:

- PTB **tuatua* 'opposite-sex sibling' (François 2005: 500)
- PTB **tuaya* 'older same-sex sibling' (Clark 2009: 203)
  < POc **tuqaka* (Pawley and Ross 2006: 53)
- PTB **tasi* 'younger same-sex sibling'
  < POc **taci* (Pawley and Ross 2006: 53).

Like Lakon, several modern languages of northern Vanuatu have preserved the protoform of their common ancestor (albeit with sound change) as well as its semantic organisation. Forms are given here for a 1sg possessor (suffix *-*gu* > -*k*):

(8) **tuatua* (-*gu*) '(my) opposite-sex sibling' >
Hiw *tʉtvɔ-k*; Lo-Toga *səse-k*; Löyöp *ʧøtʃø-k*; Mwotlap *tɪtɛ-k*; Lemerig *ʔœwœ-k*; Vera'a *ʔɔwɔ-k*; Vurës *tytyœ-k*; Mwesen *tutuɔ-k*; Mota *tutua-k*; Dorig *tutaː-k*; Koro *tɛf͡ɛa-k*; Olrat *tati-k*; Lakon *tata-k*; Mwerlap *tɞtɞ-k*.

(9) **tuaɣa* (-*gu*) '(my) older same-sex sibling' >
Vera'a *ʔuwa-k*; Vurës *toɣœ-k*;[11] Mota *tuaɣa-k*; Nume *tua-k*; Olrat *tua-k*; Lakon *tua-k*; Mwerlap *tuɔ-k*.

(10) **tasi* (-*gu*) '(my) younger same-sex sibling' >
Vera'a *ʔisi-k*; Vurës *tɪsi-k*; Mota *tasi-k*; Nume *tisi-k*; Olrat *tisi-k*; Lakon *tahi-k*; Mwerlap *tɛsi-k*.

All the forms cited in (8)–(10) have preserved the ternary system of their ancestor – the one illustrated in Figure 3 for Lakon. But the rest of the Torres–Banks languages have gone through a couple of innovations, which I will present now.

Semantic change and restructuring

The first innovation is the generalisation of **tasi* to refer to any 'same-sex sibling' regardless of relative age. This is a new example of a lexical merger (Section 3.2.1). As a corollary, the root **tuaɣa* was eliminated:

(11) **tasi* (-*gu*) * '(my) younger same-sex sibling' > '(my) same-sex sibling'
Mwotlap *ithi-k*; Lemerig *ʔisi-k*; Mwesen *tisi-k*; Dorig *tsi-k*; Koro *tsi-k*.

The five languages listed in (11) have lost the feature of *relative age* as a structuring principle in their sibling domain. The only relevant criterion is now *relative sex*: these languages contrast only two terms for siblings, namely **tuatua* 'opposite-sex sibling' vs. **tasi* 'same-sex sibling'. This innovation likely reflects a cognitive pressure for symmetry in the system.

Finally, the four northernmost languages in the area show evidence of the same change (loss of **tuaɣa*), but have added to it a further innovation. In addition to the feature *relative sex* which provides the sibling system with its overall structuring principle, these languages split the subdomain "same-sex sibling" according to the referents' *absolute sex*. This lexical split (Section 3.2.1) was made possi-

11 See also Schnell (2011: 128) for Vera'a, and Malau (2016: 284) for Vurës.

ble by allowing a new lexeme into the kinship domain. This root **tagele* (whose original meaning must have been 'other half, counterpart')[12] became used specifically for the female instances of 'same-sex sibling' – that is, a woman's sister:

(12) **tagele* (*-gu*) * '(my) counterpart'
> 'female same-sex sibling = (woman's) sister'
Hiw *takje-k*; Lo Toga *takəle-k*; Lehali *takle-k*; Löyöp *takle-k*.

As a corollary, these four languages restricted their root **tasi* to male referents, i. e., a man's brother:

(13) **tasi* (*-gu*) * 'same-sex sibling'
> 'male same-sex sibling = (man's) brother'
[Hiw *ŋʷati-k*];[13] Lo Toga *təɣi-k*; Lehali *tihi-k*; Löyöp *ʧisi-k*.

In sum, the root **tasi* (POc **taci*) has been assigned three different meanings in the course of its evolution – from its original sense 'younger same-sex sibling' to a broader meaning (11) 'same sex sibling [= woman's sister/man's brother]', and then to an innovative gendered gloss (13) 'man's brother'.

Synthesis: A diachronic map

This complex history of sibling terms in northern Vanuatu can be synthesized in Figure 4 – a diachronic map showing the three stages of lexical change. The first innovation was a *lexical merger*; the second one was a *lexical split*.

Stage 1, the initial organisation of meaning that can be reconstructed for Proto-Torres–Banks, is preserved in seven languages (including Lakon, shown in Figure 3); the two other stages illustrate successive innovations in a subset of the Torres–Banks group. Among its 17 members, ten languages went through the lexical merger of Stage 2, losing the dimension of relative age. Among these, four languages later went through the lexical split of Stage 3, as they divided the 'same-sex sibling' category according to the referent's absolute sex.

While the history of the sibling system in the region can appear complex to describe, it becomes much easier to conceive through the use of a map like Figure 4. Over time, the semantic range of each root has redesigned its contour on the map.

12 The root **tagele* is also reflected, for example, by Volow *tɛⁿgɛl X* 'across X, opposite X'; or by Mwesen *taklɛ* 'portion, part of; some; several X'.

13 Hiw has an innovative form *ŋʷati* ('♂ B') of unknown origin.

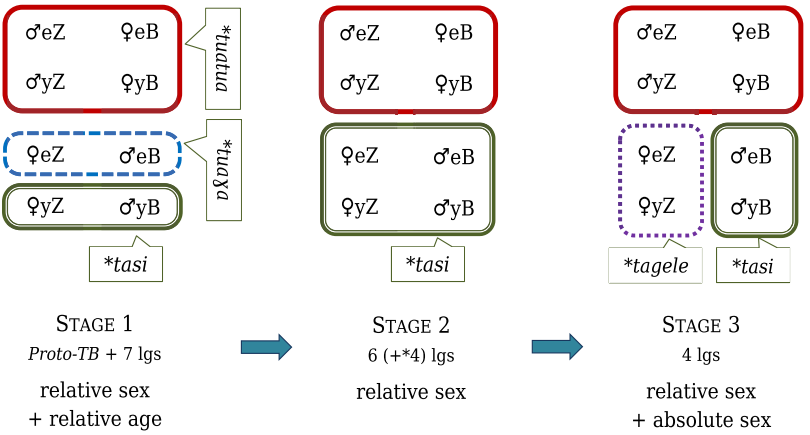


Figure 4: Terms for siblings in northern Vanuatu: three stages of evolution.

4 Lexical competition and replacement

Through examples taken from body parts or kinship terminology, we saw that a given lexical field can be characterized by structural stability (Section 3.1) or by semantic innovations (Section 3.2); these innovations may take the form of lexical mergers or lexical splits.[14]

These structural changes result in new patterns of lexification, as each language's emic categories are redefined. They can be usefully represented on a set of semantic maps, displaying the historical evolution of lexification patterns; such maps provide a clear view of which areas of the domain have remained stable, and which ones have evolved over time.

Now, a sequence of maps gives the impression of discrete stages and abrupt changes, but we know this is not exactly what happens. This raises the question of how exactly lexical innovations emerge. How can we describe the linguistic phenomenon that is being captured by these maps? What must actually happen for semantic structures to change?

14 See Section 4.5 for further types of structural innovations.

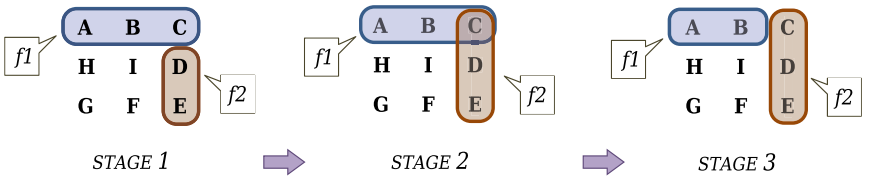


Figure 5: The three stages of lexical replacement.

4.1 Lexical replacement and the markedness shift

Based on the examples and analyses above, we can propose an abstract representation of lexical change. Let us represent a lexical domain Δ as a set of attested senses numbered ⟨A B C … H I⟩. These meanings can be spread out on a two-dimensional semantic map. Taking such a map as its background, Figure 5 illustrates three successive stages of lexical change.

In an initial Stage 1, a form *f1* colexifies together the three senses ⟨A B C⟩, whereas a form *f2* encodes two neighbouring yet distinct meanings ⟨D E⟩. Then, through a process of semantic extension – whether due to external influence (language contact) or to internal change – *f2* starts encroaching upon *f1*'s territory, as it becomes capable of also lexifying the sense C. Plotted on a map, such a change translates visually through the reshaping of each line – i. e. the blob encompassing all the senses of a polysemous word.

Stage 2 represents a phase of *lexical competition* between the words *f1* and *f2* when expressing the sense C. (By way of illustration, consider how some varieties of English show variation between *film* and *movie*, between *tap* and *faucet*, between *if it were* and *if it was*.) The variation can be a matter of stylistic register, where typically *f1* goes from being the norm to becoming marked and archaic, while *f2*, which started as a marked synonym for sense C, ends up becoming the normal, unmarked form to express it. This reversal process is called markedness shift (Dik 1989: 44); it can be captured by the formula in (14), where 'S' refers to a specific sense. (The forms in bold constitute the default terms for a given meaning, while brackets flag the synonyms that are stylistically or pragmatically "marked".)

(14)

S: ***f1***	→	S: ***f1***~(*f2*)	→	S: (*f1*)~***f2***	→	S: ***f2***
Stage 1		Stage 2a		Stage 2b		Stage 3

If the new usage takes root and spreads through the speech community (via social diffusion), eventually the old term becomes obsolete, and ends up being replaced by the new term. The final result, shown here as Stage 3, is one where sense C is

not lexified by *f1* anymore, but exclusively by *f2*. As far as this sense is concerned, the language has gone through the whole process of lexical replacement.

As Figure 5 shows, the process of lexical replacement affecting the sense C can technically be broken into two subphases:

> [Stage 2] a (partial) merger {C≠D=E} → {C=D=E}
> [Stage 3] a split {A=B=C} → {A=B≠C}

Lexical replacement, in principle, always consists of a lexical *merger* (resulting in lexical competition) followed by a lexical *split* (resolving the lexical competition).

4.2 Lexical competition as a slow, emergent process

In the reality of linguistic use, the sort of change illustrated here would take the form of a continuous process of variation between two strategies *f1* and *f2* to express a given sense (see Traugott and Dasher 2001: 11–12; Enfield 2003: 29): variation over time, variation across speakers, and even intra-speaker.

The lexical competition between variants can span over several generations, sometimes centuries. Analyses based on diachronic corpora, when they are available, show that the process of language change manifests itself as a slow evolution in the distribution of the two variants, such that *f1* remains the preferred option for a long period, while *f2* increases its prevalence over time. Rather than being a sudden shift, the emergence of *f2* is incremental; and what is transmitted from one generation to the next is really a particular statistical distribution of *f1* vs. *f2*, together with a sense of the directionality of change – as *f1* decreases towards obsolescence while *f2* grows in frequency, and evolves towards becoming the new norm. Rather than forming a linear, steady increase, the trajectory of the distribution along the time axis typically takes the form of an "S curve" (Chambers 2002: 361; Blythe and Croft 2012). The competition between the two variants begins slowly, then accelerates for a short while, before it finally fades off during another long period. This trajectory, drawn along a time axis, is visually reminiscent of the shape of the letter S or $\int$ (hence its name 'S curve', or 'sigmoid curve').[15]

This point should be kept in mind whenever a linguistic innovation (whether phonological, syntactic or lexical) is represented by an abstract formula, of the type {*f1* → *f2*} "*f1* was replaced by *f2*". Just like geological activity, lexical tectonics must be understood as a slow process, which can span over many generations and still be imperceptibly active. Competition in the lexicon involves long periods

15 While our discussion is about language change, the concept of S-curve was initially defined (Rogers 1962) in more general terms, and has been applied to various forms of social change.

when the two variants coexist in discourse. During that transitional period, the perception of norms and markedness fluctuates among speakers, before it finally stabilises.

4.3 An example of lexical replacement

To take a classic example, Late Latin progressively replaced its inherited term *caput* 'head' with another noun *testa*, originally 'earthen pot' (cf. Blank and Koch 1999). The four stages followed by this case of lexical replacement – in line with the formula in (14) – can be shown in Table 6.

The first innovation (Stage 1→2a) was to use the noun 'pot' as a figurative, slang word for 'head', in competition with the inherited form *caput* – in a way similar to early 20th c. French slang *carafe* 'jug' used sometimes for 'head'. Eventually, *testa* lost its jocular connotations, and ended up as the standard term for this body part (Stage 2b), as evidenced by Italian *testa*, and French *tête*. The older term *caput* (> Italian *capo*, O.Fr. *chief*) resisted for some time, but eventually became the marked term in the pair.

Standard Italian still belongs to Stage 2b. While *testa* has become the default name of the body part, *capo* still survives as a marked synonym – i. e. archaic, regional or medical – and in idioms (e. g., *mal di capo* 'headache'). Yet the word's main modern meaning is 'leader, chief' (e. g., *capo dello stato* 'head of State').

French has gone one step further. In Classical French, *chef* was still used as an archaic synonym of *teste*, until it disappeared in this sense. In modern French, if one puts aside a couple of vestigial compounds (e. g., *couvre-chef* 'hat'), the only living meaning of *chef* is 'leader, chief': the language has reached Stage 3, and the lexical split is now complete.

If a typologist wanted to list the languages that colexify the meanings 'head' and 'leader', Standard Italian would still qualify (with *capo*), but modern French

Table 6: When markedness shift drives lexical replacement: words for 'head' in Late Latin / Romance.

	'head'	'leader, chief'
Stage 1	*caput*	*caput*
Stage 2a	***caput*** ~ (*testa*)	*caput*
Stage 2b	(*caput*) ~ ***testa***	*caput*
Stage 3	*testa*	*caput*

would not: *le chef* is no longer used to refer to the head, and *la tête* can hardly refer to a chief.[16]

One could summarise the whole process by stating that French "replaced" its noun *caput* (Stage 1) with an innovative form *testa* (Stage 3). While not untrue, this simplified formula encapsulates what was really a gradual sequence involving lexical competition and markedness shift between polysemous terms (see Sweetser 1990: 9; Evans and Wilkins 2000: 549), spanning – in this case – over almost two millennia. Rather than a leap from Stage 1 to Stage 3, the key turning point in this evolution was really the shift from Stage 2a to Stage 2b, and the reversal of markedness.

4.4 Subtypes of lexical replacement

The process of lexical replacement, which we just examined from the angle of a single sense, can be revisited in the broader perspective of its impact upon the system's lexical structures. As a language goes through lexical replacement, a corollary is a new configuration of its lexical patterns.

As far as the time axis is concerned, we just saw that lexical replacement can be complete (cf. **chef* → *tête* in French for 'head') or it can take the form of an ongoing competition (cf. Italian *capo* ~ *testa*). This contrast can be visually represented on a map by the presence or absence of an overlap: compare Stage 2 and Stage 3 in Figure 5.

On a different dimension – namely, the distribution of meaning on the semantic map – lexical replacement can be *local* or *total*:

- *Local replacement*: *f2* replaces *f1* for only some of its senses.
- *Total replacement*: *f2* takes over *f1* in all of its senses, resulting in the elimination of *f1* from the lexicon.

For example, Table 6 showed that Lat. *caput* 'head; leader' has been replaced by *testa* only for the body part, while the sense 'leader' remains lexified by reflexes of *caput*: this is a case of *local replacement* (ongoing in Italian, complete in French).

16 Following a cyclical evolution, the analogy {leader ⇔ head} (cf. Eng. *head of state*, *head of department*) is being reactivated in modern French – this time based on *tête*, the standard term for the body part. Yet this is still limited to a few idioms, e. g., *tête de liste* 'first name on a party-list ballot'; *il est à la tête du pays* 'he is [at] the head of the country'. The word *tête* alone cannot refer to a human (**La tête viendra demain* ~ ?*The head will come tomorrow*).

Conversely, we saw in Section 3.2.1 (Figure 2) that POc **lima* 'arm, hand' was replaced by **bani* 'wing' for all its senses; this resulted in the total replacement – or *relexification* – of **lima* by **bani*, and the loss of **lima*.

4.5 A typology of innovations in patterns of lexification

Let us recapitulate the different types of innovation that can affect lexical structures.

In line with Figure 5, let us posit two word-forms *f1* and *f2*, and a segment of the semantic space defined by three senses A B C. In *stage* 1, senses {A B} are lexified by the form *f1*, and {C} by *f2*. Assuming these words undergo semantic change, we can characterise Stage 2 by how this initial pattern of lexification is modified over time. Table 7 defines five main types of diachronic scenarios: lexical competition, lexical split, lexical merger, lexical shift, and relexification.

In Table 7, the sign '—' means that the form in question no longer exists in the subdomain defined by senses A B C: either it moves on to take up a separate meaning D, or it disappears altogether from the language.

Here is a short description for each innovation type:

- *Lexical competition*:
  Sense B is initially lexified by a word *f1*; but another word *f2*, by semantic extension, acquires that same sense B. The two words *f1* and *f2* compete for some time with respect to lexifying B.
- *Lexical split*:
  A word *f1* initially colexifies two senses {A B}; but another word *f2* acquires the sense B, and after a period of lexical competition, *f2* ends up replacing *f1* for sense B. As a result, the initial pattern of colexification {A B} is split apart.

Table 7: Types of innovation in lexification patterns. Three senses A B C are distributed across word forms *f1–f2*, in different ways in stage 1 vs. stage 2.

type of innovation	Stage 1		→	Stage 2	
lexical competition	*f1*: **A B**	*f2*: **C**	→	*f1*: **A B**	*f2*: **B C**
lexical split	*f1*: **A B**	*f2*: **C**	→	*f1*: **A**	*f2*: **B C**
lexical merger	*f1*: **A B**	*f2*: **C**	→	*f1*: **A B C**	*f2*: —
lexical shift	*f1*: **A B**	*f2*: **C**	→	*f1*: **B C**	*f2*: —
relexification	*f1*: **A B**	*f2*: **C**	→	*f1*: —	*f2*: **A B**

- *Lexical merger*:
  Senses A B C used to be dislexified across two separate forms *f1* and *f2*. Following lexical replacement of *f2* by *f1* (for sense C), these senses become colexified.
- *Lexical shift*:
  A word *f1* loses some senses (e. g., A) and gains others (e. g., C) which used to be lexified with another word *f2*. Such a lexical shift combines a split ({A=B} → {A≠B}) and a merger ({B≠C} → {B=C}).
- *Relexification*:
  Senses A B used to be lexified by a word *f1*. Following wholesale lexical replacement, they become lexified by another word *f2*, while *f1* exits the domain.

5 From lexical splits to relexification

I will end this study by providing a few more examples of lexical innovations, taken from my firsthand data on Vanuatu languages. In each case, I will describe the type of innovation at stake according to the typology in Table 7. Quite often, the most powerful way to represent the semantic change in question takes the form of a dynamic semantic map, of the type that was presented in this study.

These examples will provide us with an opportunity to observe a phenomenon relatively widespread in the Vanuatu area: *relexification through semantic realignment*. Indeed, while semantic innovations tend to disrupt existing patterns of lexification (e. g., by bringing about lexical splits and mergers), languages in sustained contact show a marked tendency to later *realign* their semantic structures with those of their neighbours. It's as though the innovative word, newly introduced, underwent the pressure towards conforming its semantic outline to the general patterns of lexification that are dominant in the region. This is how lexical structures can remain stable over time, in spite of the renewal of lexical forms (see Section 3.1).

5.1 Moon, month

The Proto-Oceanic (POc) etymon **pulan* colexified {moon=month}, and most languages in northern Vanuatu have preserved that colexification.[17]

17 The CLiCS³ database [https://clics.clld.org/edges/1313-1370] lists 327 languages colexifying 'moon' and 'month'.

Table 8: 'Moon' and 'month' in a sample of northern Vanuatu languages.

	etymon	MTP	VRA	VRS	MTA	NUM	DRG	LKN	MRL
'moon'	**pulan*	nʊ-wʊl	fulʊ	wʊl	fula	wisŋar	sŋar	siŋaː	vʊl
'month'	**pulan*	nʊ-wʊl	fulʊ	wʊl	fula	wisŋar	sŋar	vʊl	vʊl

The language Lakon keeps its reflex of **pulan*, namely /vʊl/, for the sense 'month'; yet it now lexifies the moon using an innovative form /siŋaː/, from PTB **siŋaRi* 'shine' (François 2011b: 186). This is a neat case of a *lexical split* – see Table 8. In this instance, Lakon shows structural divergence compared to its linguistic neighbours.

The lexical innovation of Lakon is also attested in the other languages of Gaua, illustrated here by Dorig and Nume. Yet interestingly, these languages went one step further. They generalised the innovative root **siŋaRi* for the two meanings 'moon' and 'month', thereby eliminating **pulan* altogether from the language: e. g. Dorig now has /sŋar/ 'moon, month'. In other terms, after diverging from the other Banks languages in a way similar to Lakon, the two languages Nume and Dorig eventually "reconverged" towards the colexification pattern {moon=month} that used to be associated with **pulan*.

The result of this two-step process (divergence + re-convergence) was wholesale relexification. The polysemy {moon = month}, which had once been split apart, was eventually retrieved, but associated with a different etymon – see Figure 6.

But how can we account for such a process of semantic realignment, whereby lexicons would first lose the lexical structures of their ancestors, and then myste-

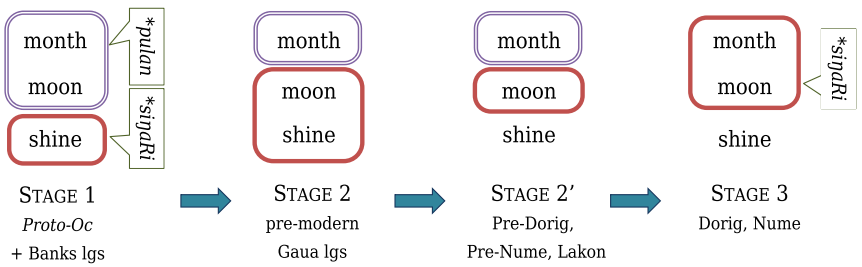


Figure 6: Language-internal relexification via semantic realignment. In Gaua languages, the colexified pair {moon = month} underwent a lexical split; then both senses were reunited through a process of contact-induced realignment.

riously retrieve them? Why didn't Nume and Dorig just keep their new dislexification {moon ≠ month} just like Lakon did? I believe the process must be explained as an effect of areal pressure. In a region characterised by widespread multilingualism (François 2012), the islanders of Gaua remain familiar with the lexical structures that are common in the Banks archipelago. And indeed, a general phenomenon observed in north Vanuatu is one of perpetual alignment of structures (whether syntactic or semantic), regardless of the form of words (François 2011a).

In that part of the world, when a language goes through a local case of lexical replacement, the resulting structural disruption usually does not last long. Once a newly introduced word has started an incursion into a given semantic domain, it is just a matter of time before it becomes harnessed into the region's dominant lexification patterns. As if by a sleight-of-hand, the end result is a complete renewal of word forms, and yet a perfect preservation of inherited polysemies. This paradox (replacement of forms, preservation of ancestral lexical structures) is what I call *language-internal relexification*.

An examination of the lexicons of north Vanuatu languages shows that language-internal relexification is there a pervasive phenomenon (François 2010; 2011a: 226–8). The typical scenario is for lexical replacement [stage 2] to be followed by semantic realignment [stage 3], just like we saw for Dorig in Figure 6. Admittedly, lexical innovation does sometimes bring about structural disruption (as in the case of Lakon here) yet that situation happens much less often. When the lexical patterns of a language are at odds with its neighbours, the units of one system become less readily translatable from one language to another; such cross-linguistic discrepancies tend to become ironed out sooner or later, so as to facilitate the handling of several lexicons in the minds of multilingual speakers.[18]

5.2 Village, island, country

A similar example is provided by the words for 'village'. The Proto-Oceanic etymon **panua* 'inhabited territory' (Pawley 2005) is well reflected in Torres–Banks languages, where it commonly colexifies {village = island = country}.

18 The constant structural convergence among north Vanuatu languages (analysed in François 2011a) is reminiscent of similar cases observed in other parts of the world, including among unrelated languages. For example, Gumperz and Wilson (1971) explained how the three languages spoken in Kupwar in India (Kannada, Marathi, Urdu) had achieved an "extraordinary degree of translatability from one local utterance to the other"; they showed how the push towards perfect translatability had been driving structural convergence among languages in contact.

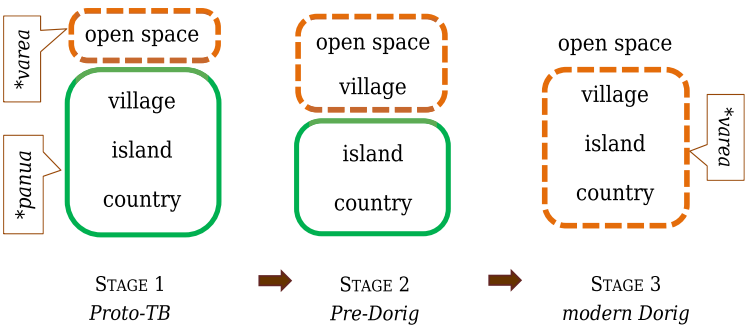


Figure 7: Language-internal relexification via semantic realignment. In Dorig, the root *varea entered the territory of *panua via the sense 'village', but eventually replaced *panua in its whole polysemy 'village, island, country'.

In Hiw, the regular reflex of **panua*, namely /vənjə/, had to concede some ground to an innovative compound /mətevənjə/ (etym. 'island spot') for the sole meaning 'village'. As a result, Hiw now splits the semantic array of **panua* across two lexical items: /vənjə/ 'inhabited territory, country, island, (*village)' vs. /mətevənjə/ 'village'. This is a case of *lexical split*.

Further south, the language Dorig has also innovated a new word for 'village', from a root **varea* 'outside, public space' (cf. Clark 2009: 218). This innovative noun must have entered the territory of **panua* through the sense 'open space in the middle of the village' → 'village'. In principle, this could have resulted in some form of lexical split (like in Hiw), with a new contrast (**varea* 'village' vs. **panua* 'island, country').

Yet what happened here, once again, was a process of semantic realignment. The lexification pattern {village = island = country} is so entrenched in the region that the new root **varea* ended up covering the whole territory of former **panua* (Figure 7). Modern Dorig has lost all reflexes of **panua*, and replaced them with a single noun /vrɛ/ 'village; island; country'. This is another perfect case of language-internal relexification.

5.3 Sacred, forbidden, haunted, cemetery

The Proto-Oceanic etymon **tabu* – the origin of Eng. *taboo* – was polysemous. It included such senses as 'sacred, holy; haunted by spiritual forces; forbidden, secret; ban, proscription; reserved to initiated men' (François 2022).

In Mwotlap, the semantic territory of **tabu* has shrunk considerably. This root only survives in the noun /nɛ.tɛk͡pʷ/ 'graveyard, cemetery' (François 2013: 225). While arguably related to the sense 'haunted place', that meaning 'graveyard' is itself innovative, as it did not even belong to the root's meanings in Proto-Oceanic.

As for the root's original polysemy, it has now been split across two roots. An etymon **salaɣoro* (literally 'forbidden path') has replaced **tabu* for all the senses reflecting a social proscription {'forbidden, secret; ban, proscription; secret society, reserved to initiated men'}. As for the spiritual meanings of **tabu*, which pertained to the notion of holiness or divine presence, they were relexified with a separate root **roŋo* ('sacred, holy, numinous' – originally 'quiet, silent').

In this case, there was no case of wholesale relexification, nor preservation of earlier lexical structures. On the contrary, the history of **tabu* in Mwotlap reflects a rich history of semantic shifts, splits, and relexifications. The root **tabu* itself has now migrated to a new location of the semantic space. As for the initial polysemy it used to have in Proto-Oceanic, it has been split apart across two other roots (**salaɣoro* and **roŋo*), resulting in quite different patterns of lexification between the modern language and its POc ancestor.[19]

5.4 Up, uphill, inland, southeast

In the domain of space directionals (François 2004; 2015), the inherited word **sake* 'up' reconstructs (for Proto Torres–Banks) as encoding a number of different directions: {⟨1⟩ [vertical] up; ⟨2⟩ [on hilly islands] uphill; ⟨3⟩ [on flat islands] inland; ⟨4⟩ [on sea] landwards'; ⟨5⟩ upwind > [inter island] southeast; ⟨6⟩ [on land] parallel to the shore, towards south}.

Now, the language Hiw has split the lexical domain of **sake* in two. An innovative directional **vene* ('climb → [go] up') has evicted **sake* – but only for the meanings most clearly linked to the vertical dimension, namely ⟨1⟩–⟨2⟩. As for senses ⟨3⟩–⟨4⟩–⟨5⟩–⟨6⟩, they are now lexified by a vestigial form /aɣ/ (François 2015: 184–185). Even though it reflects **sake* '[go] up', this form has now lost any semantic connection with verticality. The historical outcome is a neat lexical split, represented in Figure 8.

The change resulted in the dislexification between 'inland' (on gentle slopes) and 'uphill' (on steeper ground) – a rare pattern among the geocentric systems of north Vanuatu (François 2015: 151, 176). In this particular case, there was no process of semantic realignment. And indeed, the lexification patterns of space directionals in modern Hiw are quite original compared to its neighbours.

19 The reader is referred to the semantic maps of **tabu* presented in François (2022).

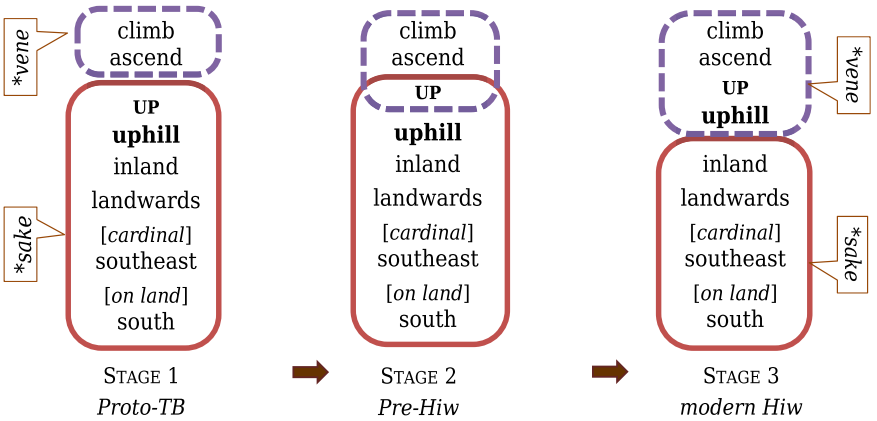


Figure 8: A lexical split. In Hiw, the original polysemy of **sake* 'up+' was split in two. Some senses are now lexified by **sake*, others by **vene*.

5.5 Give, take, get, understand

One last example concerns the polysemy of 'take', and the wholesale relexification that took place in the language Lehali.

The Proto-Oceanic root **alap-i* 'get, take' (Ross 2016: 426–427) is reflected in most of the Banks languages, via a root **lavi* or **lai*. That verb means 'take, carry'; combined with directionals, it encodes meanings such as 'bring' (take hither), 'take away', 'raise', 'lower', 'remove', and so on. Besides these basic meanings, reflexes of **lavi* can also mean 'harvest', 'steal', 'abduct', 'adopt [a child]', 'marry [a woman]'. The same verb can also take up cognitive meanings, such as 'commit s.th. to memory' or 'understand'.[20]

Finally, the combination of **lavi* 'take' with a dative argument or a participant-oriented directional (François 2003: 413, 2015: 146) forms the equivalent of 'give': 'he gave it to me' literally reads *he took it hither*. In other words, Northern Vanuatu languages are characterised by the colexification {take=give}, a pattern that is otherwise found mostly in languages of New Guinea (Gil 2017: 80).

Now one language, Lehali, has lost **lavi*, and instead uses a non-related form ***hö*** /hʊ/. This verb *hö* is cognate with Mwotlap /hɔɯ/ 'hold out o.'s arm; hand out,

20 The semantic shift from a verb of prehension 'take, get, grasp' to a cognition meaning 'understand' is common in the world (Vanhove 2008b: 361–365). Compare Lat. *capere* 'hold' > Ital. *capire* 'understand'; Fr. *saisir* 'hold, grasp' → *j'ai pas saisi ce que tu as dit* 'I didn't get what you said'.

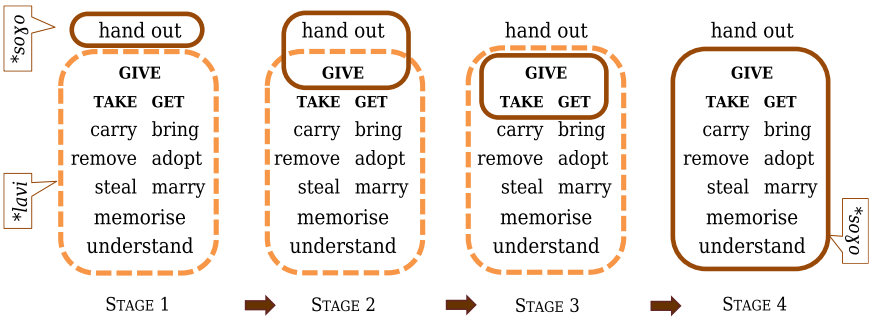


Figure 9: Language-internal relexification via semantic realignment. In Lehali, the polysemous root **lavi* 'take+' was entirely replaced by **soɣo* (etym. 'hand out').

give', Vurës /sʊɯ/ 'give generously', Mota /soɣo/ 'bring, give'. We can reconstruct a protoform for Proto Torres–Banks, **soɣo* 'hand out, give'.

Within the whole semantic set of **lavi*, the sense 'give' was probably the entry point whence **soɣo* launched its lexical intrusion, and where the competition must have begun. That sense 'give' was then to act as the "Trojan horse" in the relexification of **lavi* by **soɣo* – see Figure 9.

Figure 9 illustrates a full process of internal relexification. The word *hö* (< **soɣo*) has now replaced **lavi* in all of its senses and combinations: compare the semantic set of **soɣo* in Stage 4 with that of **lavi* in Stage 1. Besides the sense 'give', *hö* means 'take, get' in modern Lehali; it forms compounds meaning 'bring' (*hö ma*), 'remove' (*hö yak*), etc. The same verb can mean 'steal', 'adopt [a child]', 'marry [a woman]', 'understand'...:

(15) LHI *Mätēl m-van* ***hö*** *ke van l-eñ gom.*
1EX:TRI PFT-go take 3SG DIREC LOC-house sick
'We took her to the dispensary.'

(16) LHI *Ke man* ***hö*** *n-lokven, si toqo?*
3SG CPLT take ART-woman or not.yet
'Has he already taken wife, or not yet?'

(17) LHI *Nö man* ***hö***.
1SG CPLT take
'I got that.' (= I understand)

Through the process of language-internal relexification, a root whose original meaning was 'hold out o.'s arm, hand out' has reshaped its semantic array so as to include such diverse senses as 'carry', 'marry', or 'understand'.

In doing so, it appears as though the root **soɣo* had cast itself into the mould of **lavi* – like a lake that borrows its shape from the crater of an extinct volcano.

The structural properties of Lehali *hö* are now perfectly parallel with those of Mwotlap *lep*, in spite of their distinct etymology. And indeed, bilingual speakers can sometimes be heard commenting that "whenever Mwotlap says *lep* [lɛp], Lehali says *hö* [hɒ]".

* * *

In all the cases we reviewed, sequential maps proved helpful in visualising the process of lexical change as a whole.

To take a geological analogy, Figure 9 could be described as a collision between two tectonic plates: as the **soɣo* plate expanded its semantic range, it must have gone through a period of friction at the boundary, in the form of lexical competition (Section 4.1) between **soɣo* and **lavi* for the meaning 'give' (*stage* 2). Eventually, the boundary friction was resolved, so to speak, by the "subduction" of **lavi*, giving way to **soɣo* for that particular zone. As a result of these tectonic movements, the two plates **soɣo* and **lavi* changed their shapes, and redesigned the local semantic landscape.

6 Conclusion

The present study aimed at defining theoretical tools for the analysis of semantic change in the lexicon. Whether we compare languages endowed with a philological tradition like Romance, or we rely on the comparative method as is done for Vanuatu languages, semantic change can often be modelled in terms of patterns of lexification – that is, a change in distribution between forms and senses.

Many examples of lexical change can be described using a limited set of concepts: co- vs. dis-lexification; lexical merger, lexical split, relexification, semantic realignment. I also showed how these examples of structural change can be fruitfully projected onto semantic maps: these spatial representations help visualize in a powerful way the "lexical tectonics" at play in languages, as words slowly change their meanings over time.

In a way, the semantic structures of languages live a life of their own, independent of the phonological material they attach to. Structural patterns evolve in time, they compete with each other, they expand and shrink, they spread across dialect networks, migrate from language to language through calquing, and can

be shared across entire linguistic areas. Some patterns are universal, others are local and rare; they can be inherited or innovative, stable or fragile. Lexical competition between two words – itself the trigger for lexical change – may be resolved in a generation, but it may also last for centuries in the form of stylistic variation between competing synonyms.

Historical lexicology, approached through the study of lexification patterns, offers promising avenues of research in lexical typology. It also raises many questions, which can feed further research.

- Is it possible to identify some regularities in structural change? Are some innovations directional in nature (e. g., a word can change its sense from s1 to s2 but never the opposite)?
- Are lexical splits or mergers more likely to take place in certain semantic domains (kinship terms? verbs of motion, of cognition...)? Are they more prevalent in some stylistic registers? (e. g., technical vocabulary probably favours lexical splits; but poetry, slang, pidgins, emergent sign languages and avoidance registers (Evans 1992) might well foster mergers.)
- Do social factors play a role in maintaining vs. resolving lexical competition? Is structural change most often due to contact, or is it commonly triggered by language-internal factors?

The evolving geometry of meaning can usefully be captured using semantic maps that display the changing patterns of lexification. While this visualization is best achieved, for the time being, in the form of sequential static maps, in future years we could make the best of animation technologies, so as to emulate visually the ever-changing configurations of our lexicons.

Acknowledgment: I wish to thank the volume editors, two anonymous reviewers, as well as Siva Kalyan and Catherine Fuchs, for their comments on earlier versions of this work. This work contributes to the research program “Lexical typology across time and space”, within the strand *Typology and dynamics of linguistic systems* of the Paris-based network *Empirical Foundations of Linguistics* (LabEx EFL; IdEx Université de Paris).

Appendix. Language names

This paper sometimes refers to languages through a 3-letter code (distinct from ISO codes). These codes are listed here, and represented on Map 1 [Section 1.2].

DRG	Dorig	LTG	Lo-Toga	OLR	Olrat
HIW	Hiw	LYP	Löyöp	POC	Proto Oceanic
KRO	Koro	MRL	Mwerlap	PTB	Proto Torres–Banks
LHI	Lehali	MSN	Mwesen	VLW	Volow
LKN	Lakon	MTA	Mota	VRA	Vera'a
LMG	Lemerig	NUM	Nume	VRS	Vurës

References

Barðdal, Jóhanna, Elena Smirnova, Lotte Sommerer & Spike Gildea (eds.). 2015. *Diachronic construction grammar* (Construction Approaches to Language 18). Amsterdam: Benjamins.

Blank, Andreas & Peter Koch. 1999. Onomasiologie et étymologie cognitive: L'exemple de la *tête*. In Mario Vilela & Fatima Silva (eds.), *Actas do 1º Encontro Internacional de Linguística Cognitiva*, 49–72. Porto: Faculdade de Letras.

Blythe, Richard A. & William Croft. 2012. S-curves and the mechanisms of propagation in language change. *Language* 88(2). 269–304.

Brown, Cecil. 2013. Hand and arm. In Matthew Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/129 (27 July 2021).

Chambers, J. K. 2002. Patterns of variation including change. In J. K. Chambers, Peter Trudgill, Natalie Schilling-Estes (eds.), *The handbook of language variation and change*, 349–372. Malden, MA: Blackwell.

Clark, Ross. 2009. **Leo Tuai: A comparative lexical study of North and Central Vanuatu languages* (Pacific Linguistics 603). Canberra: Australian National University.

Coseriu, Eugenio. 1964. Pour une sémantique diachronique structurale. *Travaux de Linguistique et de Littérature* II(1), 139–186.

Dik, Simon. 1989. *The theory of Functional Grammar. Part I: The structure of the clause* (Functional Grammar Series 9). Dordrecht: Foris.

Dousset, Laurent. 2011. Understanding human relations: Kinship systems. In Nicholas Thieberger (ed.), *The Oxford handbook of linguistic fieldwork*, 209–234. Oxford: Oxford University Press.

Enfield, Nicholas. 2003. *Linguistic epidemiology: Semantics and grammar of language contact in mainland Southeast Asia*. London: Routledge-Curzon.

Evans, Nicholas. 1992. Multiple semiotic systems, hyperpolysemy, and the reconstruction of semantic change in Australian languages. In Günter Kellerman & Michael Morrissey (eds.), *Diachrony within synchrony: Language history and cognition* (Duisburger Arbeiten zur Sprach- und Kulturwissenschaft), 475–508. Bern: Peter Lang.

Evans, Nicholas. 2010. Semantic typology. In Jae Jun Song (ed.), *The Oxford handbook of linguistic typology*, 504–533. Oxford: Oxford University Press.

Evans, Nicholas & David Wilkins. 2000. In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language* 76. 546–592.

Filimonova, Elena (ed.). 2005. *Clusivity: Typology and case studies of inclusive-exclusive distinction* (Typological Studies in Language 63). Amsterdam: Benjamins.

Fillmore, Charles, Paul Kay & Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of 'let alone'. *Language* 64(3). 501–538.
François, Alexandre. 2003. Of men, hills and winds: Space directionals in Mwotlap. *Oceanic Linguistics* 42(2). 407–437.
François, Alexandre. 2004. Reconstructing the geocentric system of Proto Oceanic. *Oceanic Linguistics* 43(1). 1–32.
François, Alexandre. 2005. Unraveling the history of the vowels of seventeen northern Vanuatu languages. *Oceanic Linguistics* 44(2). 443–504.
François, Alexandre. 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In Martine Vanhove (ed.), *From polysemy to semantic change: Towards a typology of lexical semantic associations* (Studies in Language Companion Series 106), 163–215. Amsterdam: Benjamins.
François, Alexandre. 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*, 129–145. Louvain–Paris: Peeters.
François, Alexandre. 2011a. Social ecology and language history in the northern Vanuatu linkage: A tale of divergence and convergence. *Journal of Historical Linguistics* 1. 175–246.
François, Alexandre. 2011b. Where *R they all? The history and geography 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 Robert Mailhammer (ed.). *Lexical and structural etymology: Beyond word histories* (Studies in Language Change 11), 185–244. Berlin & New York: De Gruyter Mouton.
François, Alexandre. 2015. The ins and outs of up and down: Disentangling the nine geocentric space systems of Torres and Banks languages. In Alexandre François, Sébastien Lacrampe, Michael Franjieh & Stefan Schnell (eds.), *The languages of Vanuatu: Unity and diversity* (Studies in the Languages of Island Melanesia 5), 137–195. Canberra: Asia–Pacific Linguistics Open Access.
François, Alexandre. 2016. The historical morphology of personal pronouns in northern Vanuatu. *Faits de Langues* 47. 25–60.
François, Alexandre. 2019. A proposal for conversational questionnaires. In Aimée Lahaussois & Marine Vuillermet (eds.), *Methodological tools for linguistic description and typology* (Language Documentation & Conservation Special Publication 16), 155–196. Honolulu, HI: University of Hawaii.
François, Alexandre. 2022. Awesome forces and warning signs: Charting the semantic history of *tabu words in Vanuatu. *Oceanic Linguistics* 61(1).
François, Alexandre, Sébastien Lacrampe, Michael Franjieh & Stefan Schnell (eds.). 2015. *The languages of Vanuatu: Unity and diversity* (Studies in the Languages of Island Melanesia 5). Canberra: Asia–Pacific Linguistics Open Access.
François, Jacques. 2007. *Pour une cartographie de la polysémie verbale* (Collection Linguistique 92). Leuven: Peeters.
François, Jacques. 2013. Semantische Karten für die vergleichende lexikologische Analyse europäischer Sprachen. *Journal for EurolinguistiX* 10 (2013). 15–28.

Georgakopoulos, Thanasis & Stéphane Polis. 2018. The semantic map model: State of the art and future avenues for linguistic research. *Language and Linguistic Compass* 12(2). e12270. https://doi.org/10.1111/lnc3.12270.

Georgakopoulos, Thanasis & Stéphane Polis. 2021. Lexical Diachronic Semantic Maps. Mapping the evolution of time-related lexemes. *Journal of Historical Linguistics* 11(3). 367–420.

Georgakopoulos, Thanasis, Daniel A. Werning, Jörg Hartlieb, Tomoki Kitazumi, Lidewij van de Peut, Annette Sundermeyer & Gaëlle Chantrain. 2016. The meaning of ancient words for 'earth': An exercise in visualizing colexification on a semantic map. *eTopoi. Journal for Ancient Studies* 6. 418–452.

Gil, David. 2017. Roon *ve*, DO/GIVE coexpression, and language contact in Northwest New Guinea. *NUSA: Linguistic studies of languages in and around Indonesia* 62. 43–100.

Goddard, Cliff. 2001. Lexico-semantic universals: A critical overview. *Linguistic Typology* 5(1). 1–65.

Goldberg, Adele E. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences* 7(5). 219–224.

Gumperz, John J. & Robert Wilson. 1971. Convergence and creolization: A case from the Indo-Aryan/Dravidian border in India. In Dell Hymes (ed.), *Pidginization and creolization of languages*, 151–168. Cambridge: Cambridge University Press.

Hamberger, Klaus, Michael Houseman & R. White Douglas. 2011. Kinship network analysis. In John Scott & Peter J. Carrington (eds.), *The SAGE handbook of social network analysis*, 533–549 London: SAGE Publications.

Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Michael Tomasello (ed.), *The new psychology of language*, 211–243. New York: Erlbaum.

Hjelmslev, Louis. 1943. *Omkring sprogteoriens grundlæggelse* [Prolegomena to a theory of language]. København: Ejnar Munksgaard.

Juvonen, Päivi & Maria Koptjevskaja-Tamm (eds.). 2016. *The lexical typology of semantic shifts* (Cognitive Linguistics Research 58). Berlin & Boston: De Gruyter Mouton.

Koptjevskaja-Tamm, Maria & Henrik Liljegren. 2017. Semantic patterns from an areal perspective. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 204–236. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107279872.009.

Koptjevskaja-Tamm, Maria, Martine Vanhove & Peter Koch. 2007. Typological approaches to lexical semantics. *Linguistic Typology* 11(1). 159–186.

List, Johann-Mattis, Simon Greenhill, Cormac Anderson, Thomas Mayer, Tiago Tresoldi & Robert Forkel. 2018. CLICS²: An improved database of cross-linguistic colexifications assembling lexical data with the help of cross-linguistic data formats. *Linguistic Typology* 22(2). 277–306.

Malau, Catriona. 2016. *A grammar of Vurës, Vanuatu* (Pacific Linguistics 651). Berlin & Boston: De Gruyter Mouton.

Matisoff, James A. 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 (ISHI).

Osmond, Meredith & Malcolm Ross. 2016. The human body. In Malcolm Ross, Andrew Pawley & Meredith Osmond (eds.), *The lexicon of Proto-Oceanic. Vol. 5. People: Body and mind* (Asia-Pacific Linguistics 28), 75–208. Canberra: Australian National University.

Pawley, Andrew. 2005. The meaning(s) of Proto Oceanic *panua. In Claudia Gross, Harriet Lyons & Dorothy Counts (eds.), *A polymath anthropologist: Essays in honour of Ann Chowning*, 211–223. Auckland: University of Auckland, Department of Anthropology.

Pawley, Andrew & Malcolm Ross. 2006. The prehistory of Oceanic languages: A current view. In Peter Bellwood, James Fox & Darrell Tryon (eds.), *The Austronesians: Historical and comparative Perspectives*, 39–80. 2nd edn. Canberra: Australian National University.

Rakhilina, Ekaterina & Tatiana Reznikova. 2016. A frame-based methodology for lexical typology. In Pävii Juvonen & Maria Koptjevskaja-Tamm (eds.), *The lexical typology of semantic shifts* (Cognitive Linguistics Research 58), 95–129. Berlin & Boston: De Gruyter Mouton.

Rogers, Everett M. 1962. *Diffusion of innovations*. Glencoe, IL: Free Press.

Ross, Malcolm. 2016. Posture and movement. In Malcom Ross, Andrew Pawley & Meredith Osmond (eds.), *The lexicon of Proto-Oceanic. Vol. 5. People: Body and mind* (Asia-Pacific Linguistics 28), 367–456. Canberra: Australian National University.

Ross, Malcolm, Andrew Pawley & Meredith Osmond (eds.). 2016. *The lexicon of Proto-Oceanic. Vol. 5. People: Body and mind* (Asia-Pacific Linguistics 28). Canberra: Australian National University.

Schapper, Antoinette. 2021. Baring the bones: The lexico-semantic association of bone with strength in Melanesia and the study of colexification. *Linguistic Typology*. https://doi.org/10.1515/lingty-2021-2082.

Schapper, Antoinette, Lila San Roque & Rachel Hendery. 2016. Tree, firewood and fire in the languages of Sahul. In Pävii Juvonen & Maria Koptjevskaja-Tamm (eds.), *The lexical typology of semantic shifts* (Cognitive Linguistics Research 58), 355–422. Berlin & Boston: De Gruyter Mouton.

Schnell, Stefan. 2011. *A grammar of Vera'a: An Oceanic language of North Vanuatu*. Kiel: Universität Kiel dissertation.

Sweetser, Eve E. 1990. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure* (Cambridge Studies in Linguistics 54). Cambridge: Cambridge University Press.

Traugott, Elizabeth Closs & Richard B. Dasher. 2001. *Regularity in semantic change* (Cambridge Studies in Linguistics 97). Cambridge: Cambridge University Press.

Ullmann, Stephen. 1957. *The principles of semantics*. 2nd ed. Glasgow, Oxford: Jackson, Basil Blackwell.

Urban, Matthias. 2012. *Analyzability and semantic associations in referring expressions*. Leiden: Leiden University dissertation.

Vanhove, Martine (ed.). 2008a. *From polysemy to semantic change: Towards a typology of lexical semantic associations* (Studies in Language Companion Series 106). Amsterdam: Benjamins.

Vanhove, Martine. 2008b. Semantic associations between sensory modalities, prehension and mental perceptions. In Martine Vanhove (ed.), *From polysemy to semantic change: Towards a typology of lexical semantic associations* (Studies in Language Companion Series 106), 341–370. Amsterdam: Benjamins.

Wierzbicka, Anna. 1992. *Semantics, culture, and cognition: Universal human concepts in culture-specific configurations*. Oxford: Oxford University Press.

Wilkins, David. 1996. Natural tendencies of semantic change and the search for cognates. In Mark Durie & Malcolm Ross (eds.), *The comparative method reviewed*, 264–304. Oxford: Oxford University Press.