HAL
open science

Semantic maps and the typology of colexification: Intertwining polysemous networks across languages

Alexandre François

To cite this version:

Alexandre François. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. Martine Vanhove. From Polysemy to Semantic change: Towards a Typology of Lexical Semantic Associations, Benjamins, pp.163-215, 2008, Studies in Language Companion Series. halshs-00526845

HAL Id: halshs-00526845

https://shs.hal.science/halshs-00526845

Submitted on 18 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semantic maps and the Typology of Colexification

Intertwining polysemous networks across languages

Alexandre FRANÇOIS

LACITO-CNRS, Paris

francois@vjf.cnrs.fr

Abstract

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 polysemies associated with the notion "BREATHE". Intertwined together, they compose a single, universal network of potential semantic extensions.

Semantic maps and the Typology of Colexification

Intertwining polysemous networks across languages

Alexandre FRANÇOIS

LACITO-CNRS, Paris

1 General issues of lexical typology[1]

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 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 to take 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."[2] 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 facts of polysemy are stated explicitly from the perspective of a multiplicity of senses. The idea that each polysemy is fundamentally underlied 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 polysemies

A first illustration can be proposed, with the English word *straight*. Roughly speaking, this adjective may be broken into at least the following senses[3] (see 3.1 for a discussion of the method):

⟨rectilinear⟩	(*a straight line*)	⟨heterosexual⟩	(*gay or straight*)
⟨frank⟩	(*straight talking*)	⟨undiluted⟩	(*straight whisky*)
⟨honest⟩	(*a straight guy*)	⟨directly⟩	(*straight to the point*)
⟨classical⟩	(*a straight play*)	⟨immediately⟩	(*straight away*)

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

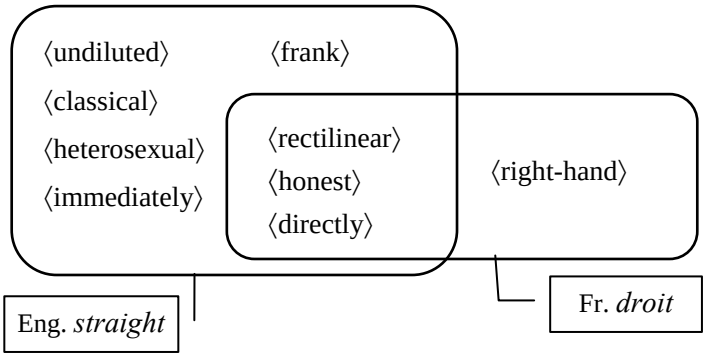
⟨rectilinear⟩	(*un trait droit*)
⟨directly⟩	(*aller droit au but*)
⟨honest⟩	(*un type droit*)
⟨right-hand⟩	(*le côté droit*)

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, 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”.

Figure 1 – Overlapping polysemies: Eng. *straight* vs Fr. *droit*



In sum, the fine-grained comparison of lexicons across the world's languages can be efficient provided each polysemous network is first broken down into its semantic atoms or “senses”. This may be done regardless of one's theoretical preferences – whether these senses are taken as actual semantic sub-categories in the speakers' minds, or merely contextual manifestations of a deeper meaning. This approach, whereby a given word is analyzed into its semantic atoms, is the first step before languages can be compared with precision, showing which senses each language lexifies together. In this new perspective, the primary unit of observation for lexical typology is no longer the *word* – a complex, highly language-specific entity – but the *sense* – a functionally-based, language-independent criterion (3.1).

These observations form the basic principles of the model I will introduce in the remainder of this article. Section 3 will first discuss the methodology for isolating senses, and for observing the way languages group them together; I will then introduce the concept of “colexification”. Section 4 will discuss the principles underlying the representation of lexical semantic maps, drawing on the principles set out by Haspelmath (2003). The model here delineated should provide empirical tools for the observation and analysis of polysemy across languages. Hopefully, it should also make it possible for future research 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 against 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 (§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.[4]

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[5]."

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?

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 favor 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 favor 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 (§5.2; 0), with the use of respectively '+' vs '[+]' signs; and in maps (§9.3), 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 may 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 s1 and s2, several configurations may come out of the data, suggesting possible questions for the typological study of the lexicon.

In case the colexification of s1 and s2 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 s1 and s2 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 s1 and s2 a case of metaphor, metonymy, hyperonymy, analogical extension…? Is it possible to reconstruct the direction taken historically by this extension (from s1 to s2, or the reverse)? Is it useful to reconstruct a missing link[6] 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 s1 and s2, 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.[7] 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 s1 and s2 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.

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 colored spots.[8] 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 leveling, 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 *sentīre* 'feel' and *audīre* '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 *roŋo 'hear, smell, feel' (Clark n.d.), from Proto Oceanic *roŋoʀ.
– 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 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⟩.[9] 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.

Another variant of this formula would be (4):

(4) "If a language colexifies s1 and s2, then it will not colexify s2 and s3."

– which may be also formulated as:

(4') "Although there may exist languages that colexify s1 and s2, and others that colexify s2 and s3, no language colexifies together the three senses s1–s2–s3."

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 s1 and s2, then it will have the linguistic property P."

(5') "If a language has the linguistic property P, then it will colexify s1 and s2."

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 s1 and s2, then its environment will have the property P."

(6') "If a language's environment has the property P, then this language will colexify s1 and s2."

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* p.6) 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,[10] 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.[11]

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.

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.[12] 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.

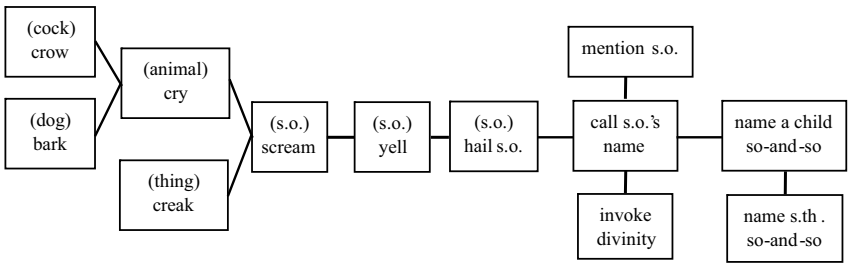
Judgments of closeness between senses are established in a dual fashion: first, by taking into account the ontological properties of each sense; second, by examining empirical data from various languages.

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.):[13] ⟨creak⟩; ⟨name a child so-and-so⟩; ⟨crow⟩; ⟨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 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 englobe 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 center. 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 s1 and s2 are attested to be colexified in the world's languages, the map centered on s1 will be a different map from the one centered on s2. 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).[14]

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 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.ḥ*), in Nahuatl (verb *imi'iyo*), in Mwotlap (verb *mōkheg*), in Nêlêmwa (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 *mōkheg* 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 Nêlêmwa verb *horêân* has added an extension to the meaning ⟨take a rest⟩, namely ⟨stop doing s.th., cease⟩ (e.g. *Co horêân o khiiboxa 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 *m̃apu* 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 Nêlêmwa, 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 finallyq 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).[15]

Table 1 – Examples of colexification associated with {BREATHE}

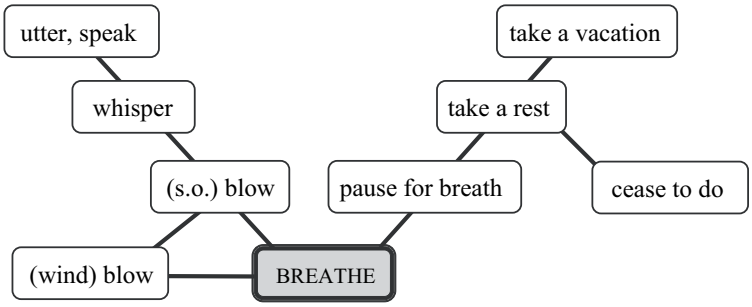
	ENGLISH *breathe*	RUSSIAN *du[x]*	MWOTLAP *mōkheg*	NÊLÊMWA *horêân*	ARAKI *soro*	FRENCH *souffler*
BREATHE	+	+	+	+		
take a rest	[+]	[+]	+	+		+
be on vacation		[+]	+			
cease to do				+		
(wind) blow	+	[+]				+
(s.o.) blow	+	[+]	+		+	+
whisper	+				+	+
utter, speak	+			[+]	+	+

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⟩.

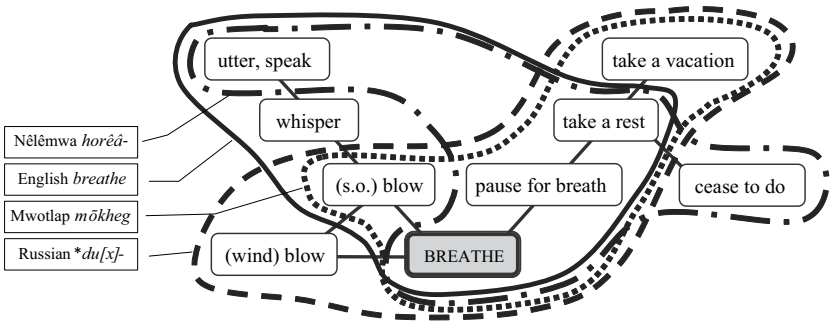
A more informative and graphic representation would thus take the form of a semantic map, a diagram showing all the senses attested, together with the most likely semantic connections that link them. These connections are first based on intrinsic semantic properties, and are then checked against empirical data (see 4.2).[16] This brings about the tentative map of *Figure 3*.

Figure 3 – A first semantic map for {BREATHE}



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 tētēnge* '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*.[17]

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. *vīta*, Arabic *ʕaiša*…), 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*[18] 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 *spīritus*, 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 *spīritus* 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 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 *spīritus*, whose semantic array is already wide (9.1.3).

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} (9.1.10): e.g. *rūḥ al-taʿāun* '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]*[19] (9.1.4).

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 behavior, 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 beings, whether in an animist or a monotheist context. This polysemy can be illustrated again with Fr. *esprit* (< *spīritus*): 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*, *spīritus*) and of its daughter languages. Surprisingly similar patterns of polysemy are found elsewhere: Greek *psūkhē* and *pneuma*; Sanskrit *ātman*; Russian *du[x]*; Arabic *rūḥ* and *nafs*; Aleut *anri*; Nahuatl *imi'iyo*; 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⟩
— ⟨immaterial 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 immateriality 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 *inspīro* 'blow into'. Likewise, the Classical Greek *pneuma*, literally 'breath, breath of air…' is attested with the meaning 'divine breath',[20] 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 (9.1.1), 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-jña* 'knowing one's self', *ātmêśvara* 'master of one's self', *ātma-ghāta* 'suicide', *ātma-grāhin* 'selfish'…

Likewise, the ordinary reflexive marker for Standard Arabic is *nafs-ī* (1sg

possessed form of *nafs*, parallel to Eng. 'my-self'). This is in fact a noun *nafs* meaning ⟨soul⟩, ⟨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.ḥ*, apparently does not go that far, at least for Standard Arabic. However, Naïm (2007: 315) reports the grammaticalization of *rūḥ* as a reflexive marker in modern Yemeni Arabic (as in *ʔalaṭṭim rūḥ-ī* 'I'm hitting myself').[21] 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*).[22]

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 p.34), 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 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 Notes

[1] 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.

[2] "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).

[3] Throughout this paper, angled brackets ⟨…⟩ are used to represent senses, insofar as they form an element of a polysemous network.

[4] 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 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.

[5] The term "lexical form" may refer to a lexeme or a construction, or occasionally to a lexical root (but see below for a discussion).

[6] 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.

[7] "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.

[8] 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).

[9] 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⟩.

[10] See, for example, Anderson (1982) for the perfect; Croft *et al.* (1987) for the middle voice; Jurafsky (1996) for the diminutive; etc.

[11] For a visual illustration of this principle, see *Figure 4* p.23, and the figures in *Appendix 3*.

[12] 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 (this volume) for French; or software such as Thinkmap's *Visual Thesaurus*® for English (www.visualthesaurus.com).

[13] 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, or the intertwining of several such networks into one – are fundamentally the same, at this particular stage of the study.

[14] 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*.

[15] The typographical contrast between *plain plus* '+' and *bracketed plus* '[+]' corresponds respectively to strict and loose colexification (see 3.2).

[16] To be precise, the polysemy of Nêlêmwa *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 Nêlêmwa 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 Nêlêmwa set appears as non-contiguous in *Figure 4*, in spite of the ideal design of semantic maps in Haspelmath's terms.

[17] 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 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).

[18] 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 9.1.3); *animus* is only included in the corpus by virtue of its synchronic cognacy with *anima* ("loose colexification").

[19] Amongst the various lexical items that are etymologically related with this root, the noun *duša* 'soul, spirit…' has received special attention in Wierzbicka (1992: 31sqq.).

[20] 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 *psūkhē* '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 *rūḥ* in the appendix). Exactly in the same way, the semantic calque took place in Latin with *spīritus* 'breath, blow of air; soul…' rather than *animus*, because the connection of the latter noun with 'wind, breath' was then no longer perceptible.

[21] Technically, although this is an instance of the noun *rūḥ*, this specific polysemy should count as loose rather than strict colexification (see 3.3) in the map of *rūḥ*, 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* p.48.

[22] A comparison between Appendix 1 and Appendix 2 shows that the universal map retains in fact only the typologically significant cases of colexification – that is, generally those that are attested at least in two languages. Some very specific senses, found in only one language (e.g., Sanskrit *ánila* 'rheumatism'; Greek *psūkhē* 'butterfly'…), have not been included, to gain space and readability. Ideally such isolated senses should be able to be included in the map, or at least kept somewhere in the matrix database – in case they turn out to be attested again in other languages when the corpus widens.

8 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-Philadelphia: John Benjamins.

Bailly, A. 1950 [1894]. *Dictionnaire Grec–Français*. Paris: Hachette.

Bril, I. 2000. *Dictionnaire nêlêmwa-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.

——. 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. Unpublished computer files: University of Auckland.

Croft, W. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

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: 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 G. Kellerman & M. Morrissey (eds), *Diachrony within Synchrony*. Bern: Peter Lang Verlag. Pp.475-508.

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: University of Alaska.

François, A. 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. [downloadable at http://alex.francois.free.fr/]

——. 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.

——. 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.

——. 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. New York: Erlbaum.

Haspelmath, M., Dryer, M., Gil, D. & Comrie B. (eds). 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.

Jurafsky, D. 1996. Universal tendencies in the semantics of the diminutive. *Language* 72, 533-578.

Kay, P. & Luisa M. 2005. Colour Terms. In Haspelmath *et al*. (eds). Pp.534-545.

Lakoff, G. 1987. *Women, Fire and Dangerous Things: What Categories reveal about the Mind*. Chicago, London: University of Chicago Press.

Langacker, R. 1987. *Foundations of Cognitive Grammar, vol.1: Theoretical prerequisites*. Stanford: Stanford University 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, vol. VI]. Philadelphia: Institute for the Study of Human Issues.

Monier-Williams, M. 1970 [1899]. *A Sanskrit-English Dictionary, Etymologically and Philologicaly Arranged with Special Reference to Cognate Indo-European Languages*. Delhi: Motilal Banarsidass.

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.
Pawley, A. & Syder, F. H. 1983. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In *Language and Communication*, J. Richards & R. Schmidt (eds), 191-225. London: Longman.
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 Beḍawi. 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), 1995.
Saussure, F. de. 1972 [1916]. *Cours de linguistique générale*. Text edited and presented by Tullio 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: Thomas Nelson.
Wierzbicka, A. 1992. *Semantics, Culture, and Cognition: Universal Human Concepts in Culture-Specific Configurations*. New York, Oxford: Oxford University Press.
Wilkins, D. 1996. Natural tendencies of semantic change and the search for cognates. In *The comparative method reviewed*, M. Durie & M. Ross (eds), 264-304. New York: Oxford University Press.

9 Appendices

9.1 Appendix 1: Lexical data

9.1.1 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):

ātma-vat animated, having a soul.
ātma- 1. soul. 2. self, one's own…
ātma-grāhin taking for one's self, selfish.
ātma-ghāta suicide.
ātma-jña 1. knowing one's self. 2. knowing the supreme spirit.
ātma-jyotis the light of the soul or supreme spirit.
ātma-tā essence, nature.
ātma-dā granting breath or life.
ātma-pāta descent of the soul, re-birth.
ātma-bhāva 1. existence of the soul. 2. the self, proper or peculiar nature.
ātma-yoga union with the supreme spirit.
ātma-víd knowing the nature of the soul or supreme spirit.
ātma-sáni granting the breath of life.
ātma-dhīna 1. depending on one's own will. 2. one whose existence depends on the breath or on the principle of animal life: sentient.
ātmê-śvara master of one's self.

Cognate form: ***an*** (3sg *án-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):

aná breath, respiration.
aná-vat-va the state of being endowed with breath or life.
anana breathing, living.
ánila 1. air, wind. 2. the god of wind. 3. wind as one of the humors of the body. 4. rheumatism.

9.1.2 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):

psūkhikos 1. vital. 2. living being; animal. 3. terrestrial, material. 4. of the soul, spiritual.
psūkhō 1. breathe, blow air. 2. breathe out, reject. 3. cool down (s.th.). 4. get cold; fall, die.
psūkhos 1. fresh breath of air; cold air, coldness. 2. winter.
psūkhros 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.

9.1.3 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):

  animal living being, animal.
  animōsus 1. courageous, bold. 2. proud. 3. ardent.

- Lexical item: ***spīritus***

  1. air in motion, breath of air. 2. act or manner of breathing; breath. 3. sigh. 4. smell, scent. 5. (breath of) life; principle of life. 6. divine breath; magic or poetic inspiration. 7. soul (opp. body). 8. self, person. 9. pride, arrogance, self-importance. 10. feelings, state of mind, moral disposition.

  Cognate form: ***spīro***.

  1. (wind) blow. 2. (sea) bubble. 3. breathe. 4. be alive. 5. be inspired. 6. (s.o., s.th.) exhale a smell. 7. breathe out (s.th.). 8. (fig) exude [cruelty…].

9.1.4 Russian

Sources: Sakhno (2005: 89 sqq.); Pauliat (1991).

- Lexical item: ***dux***.

  1. breathing, breath. 2. breath of life. 3. spirit (of s.o./s.th.): moral disposition, frame of mind. 4. mood (good or bad). 5. morale, courage. 6. supernatural being (good or evil); God (*Svjatoj dux* 'Holy Spirit'). 7. ghost.

Cognate, derived and compound forms (selection):

dut'	1. (s.o./wind) blow. 2. drink a lot.
zaduvat'	1. (wind) start blowing. 2. (s.o.) blow (candle+).
vozdux	1. air. 2. open space, outside.
dyšat'	breathe.
doxnut'	breathe, blow.
dyxanie	breathing, breath.
dyxatel'nyj	respiratory.
dušit'	strangle, suffocate; oppress.
uduš'e	breathlessness, asthma.
doxnut'	die.
izdyxat'	die.
vzdyxat'	sigh.
otdyxat'	take rest.
otdyx	1. pause, rest. 2. leisure, vacation.
peredyška	short pause, respite.
dušok	bad smell.
duxi	perfume.
duša	1. soul, spirit. 2. seat of feelings, heart. 3. inhabitant, person.
duševnyj	1. psychic, mental. 2. sincere, cordial.
duxovnyj	spiritual; holy, sacred; ecclesiastical.
duxovenstvo	clergy.
vdoxnovenie	(poetic/magic) inspiration, enthusiasm.
oduševlënnyj	animate.

9.1.5 Mandarin Chinese

Source: [no author] (1990); [no author] (1996).

- Lexical item : ***qì***.

  1. weather, atmosphere. 2. gas. 3. air. 4. (s.o.'s) breath. 5. smell (good or bad), scent. 6. (s.o.'s) manner, ways, attitude, style. 7. (s.o.'s) spirits, moral strength, morale. 8. energy, vital force, vital breath. 9. annoy, irritate (s.o.).

Cognate, derived and compound forms (selection):

qìhòu	weather, climate.
qìxiàng	1. meteorology. 2. (fig) atmosphere.
qìchuǎn	breathless, asthmatic.
qìxī	1. breath; last breath. 2. smell, scent. 3. (fig) taste, style, fashion.
qìwèi	1. smell, scent. 2. (fig) taste, style, fashion.
qìpài	style; stylishness.
qìpò	1. character strength; boldness, daring; will. 2. majesty.
qìshì	1. strength, vigor, energy. 2. momentum, impetus. 3. majesty.
qìlì	strength, vigor, energy; effort.
qìnǎo	get angry.
qìfèn	furious, angry, exasperated.

qìhuà	angry words; words (*huà*) uttered in a fit of anger.	*qìjié*	honesty, frankness.
qìshì xiōngxiōng	fierce, furious, arrogant.	*qìliàng*	open-mindedness, tolerance.
qìyàn	arrogance, insolence.	*qìněi*	disheartened, depressed.
qìgài	manner, ways.	*qìxing*	character, disposition; mood.
qìzhì	character, (good) moral qualities.	*qìshèng*	be in a bad mood.

9.1.6 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:

aniqniq	breathing; breath.	*aniqnii-q-pu-q*	breathe one's last, die.
aniqsaaqtuq-puq	breathe.	*aniqsaaq*	spirit, ghost.

  Cognate form: (Aleut) ***anr(i)*** 'breathe, blow'.
  1. breath. 2. voice. 3. principle of life, life. 4. spirit, soul. 5. ghost.
  Cognate form: *anrari* 'be alive'.

9.1.7 Nahuatl

Spoken in Mexico. Source: Marie-Noëlle Chamoux (pers. comm.)

- Lexical item: ***imi'iyo ~ i'iyac***.

  1. breath. 2. smell (esp. bodily smell). 3. *V* + *imi'iyo* 'take rest'.

9.1.8 Mwotlap

Austronesian; Oceanic subgroup. Spoken in Vanuatu. Source: François (in prep.)

- Lexical item: ***nō-mōkhe***.

  1. (s.o.'s) breath. 2. breath of life, life; principle of life. 3. 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.

9.1.9 Nêlêmwa

Austronesian; Oceanic subgroup. Spoken in New Caledonia. Source: Bril (2005)

- Lexical item: ***horêâ-t***.

  1. (s.o.'s) breath, breathing. 2. breath of life, life. 3. spoken message.

  Cognate form: ***horêân***.

  1. breathe. 2. be alive. 3. pause, take rest. 4. cease to do.

9.1.10 Standard Arabic

Source: Reig (1983); Naïm (2007).

- Lexical item: ***rūḥ*** [root *r.w.ḥ*]

  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):

rūḥī	1. spiritual. 2. alcoholic.	*istirāḥa*	take a rest, relax, be quiet.
rūḥānī	spiritual; divine; immaterial; sacred.	*rāḥah*	rest, quietness; ease, comfort.
arwāḥiyah	animism.	*rūḥ-ī*	*lit*. ‘my soul’ > grammaticalized as a reflexive marker (‘myself’) in modern Yemeni Arabic
rīḥ	breath of air, wind.		
rāʾiḥah	smell, scent (good or bad).		
mirwaḥah	fan, propeller.		
rawwaḥa	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):

tanaffasa	1. breathe, blow. 2. be reassured.	*nafsiya*	psychology.
tanaffus	breathing, breath.	*naffasa*	comfort, appease.
nafsī	psychic, psychological.	*tanāfasa*	compete, rival

9.1.11 Bedja

Afro-asiatic family. Spoken in Sudan.
Source: Martine Vanhove (pers. comm.), after Roper (1928).

- Lexical item: ***šūk***.

  1. breathing, breath. 2. soul.

9.1.12 Makonde

Niger-Congo; Bantu subgroup. Spoken in Tanzania. Source: S. Manus (pers. comm.)

- Lexical item: ***ku-pumula***.

  1. breathe. 2. take rest.

9.1.13 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’

9.2 Appendix 2: Results and semantic map

<Table 2 here>

<Figure 5 here>

see last pages of this file
(special margins for Figures and Tables
as required by Benjamins publisher
+ Landscape size)

9.3 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 languages 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 to supplement 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.

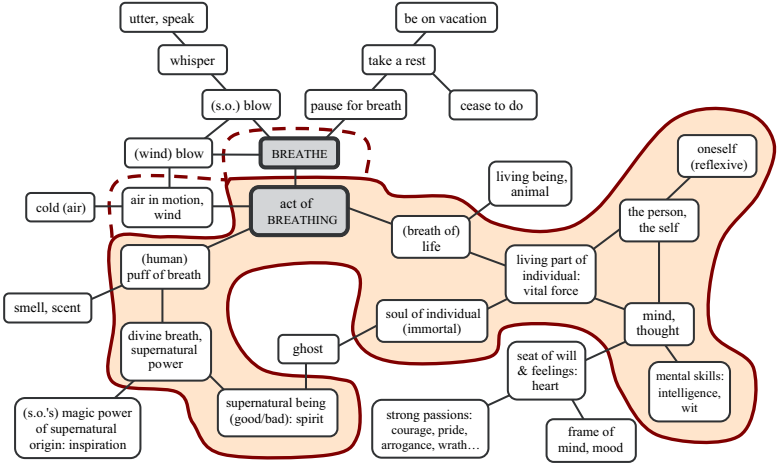


Figure 6 – Isolectic set for Sanskrit *ātman*

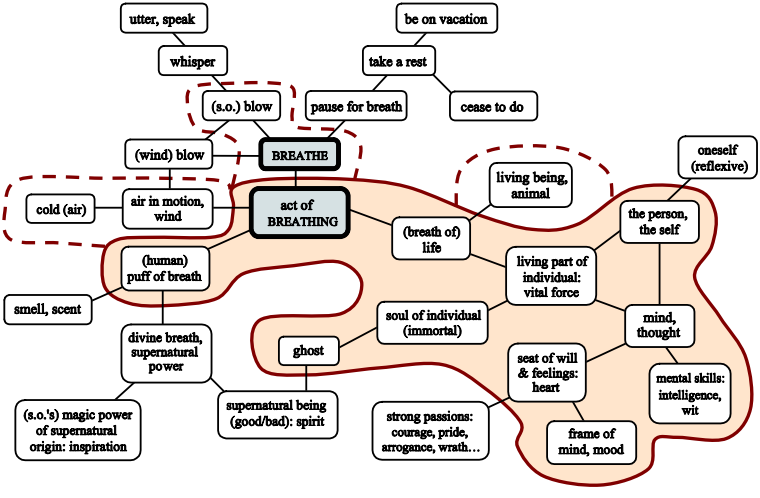


Figure 7 – Isolectic set for Greek *psūkhē*

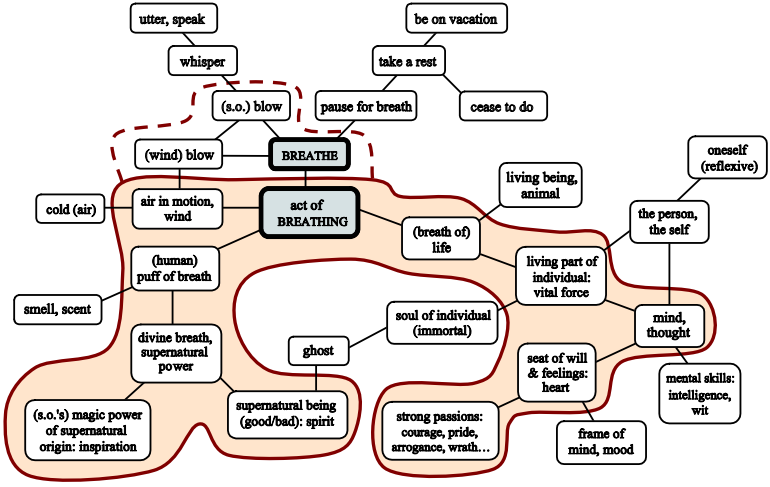


Figure 8 – Isolectic set for Greek *pneuma*

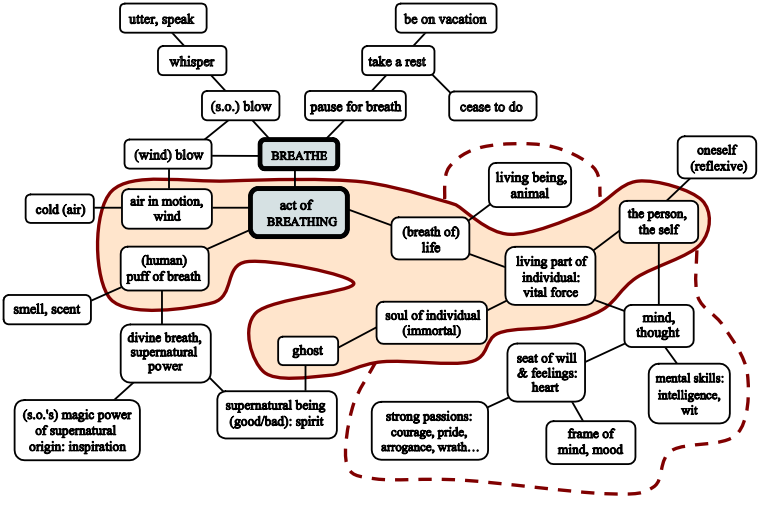


Figure 9 – Isolectic set for Latin *anima*

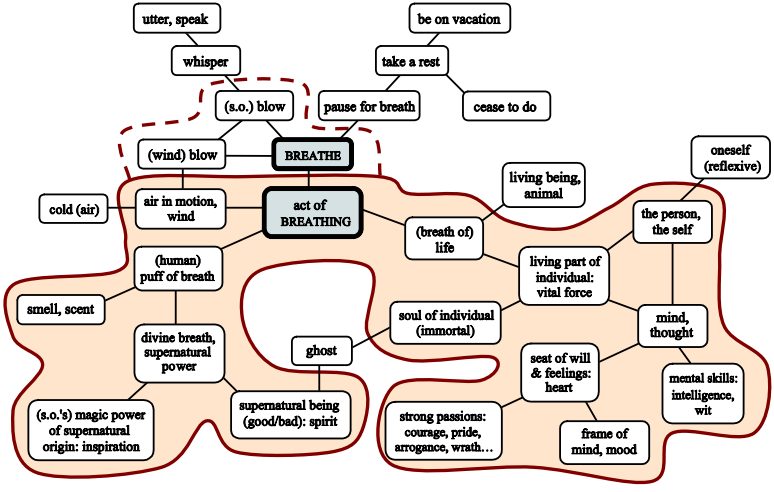


Figure 10 – Isolectic set for Latin *spīritus*

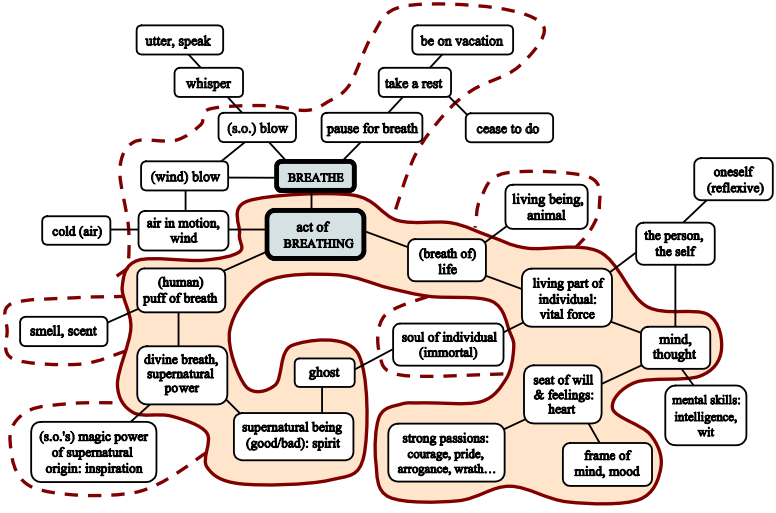


Figure 11 – Isolectic set for Russian *dux*

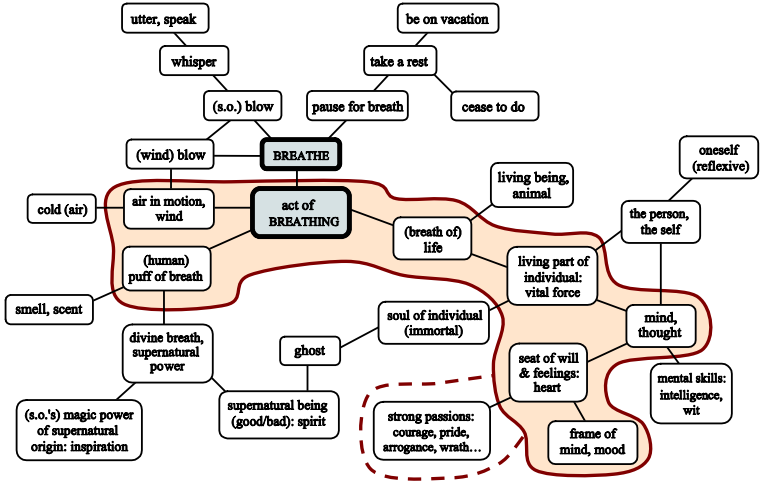


Figure 12 – Isolectic set for Chinese *qì*

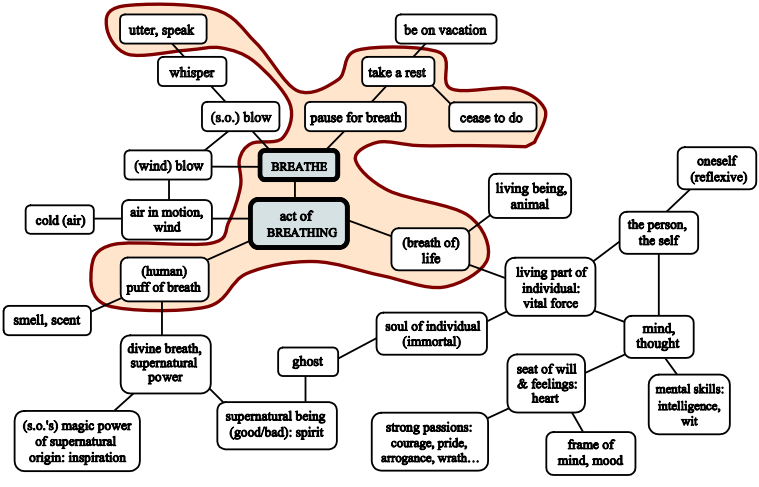


Figure 13 – Isolectic set for Nêlêmwa *horêâ-*

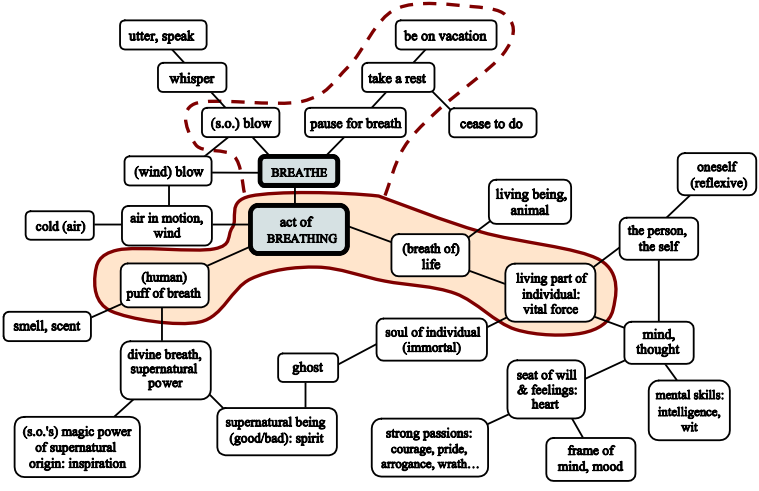


Figure 14 – Isolectic set for Mwotlap *mōkhe*–

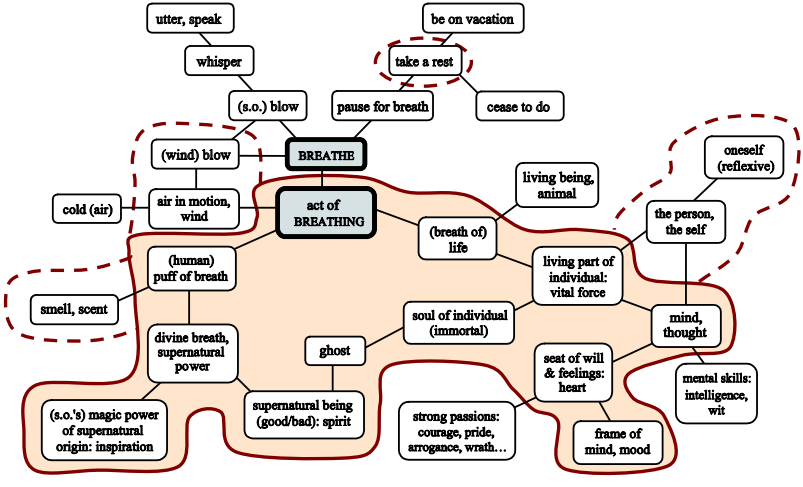


Figure 15 – Isolectic set for Standard Arabic *rūḥ*

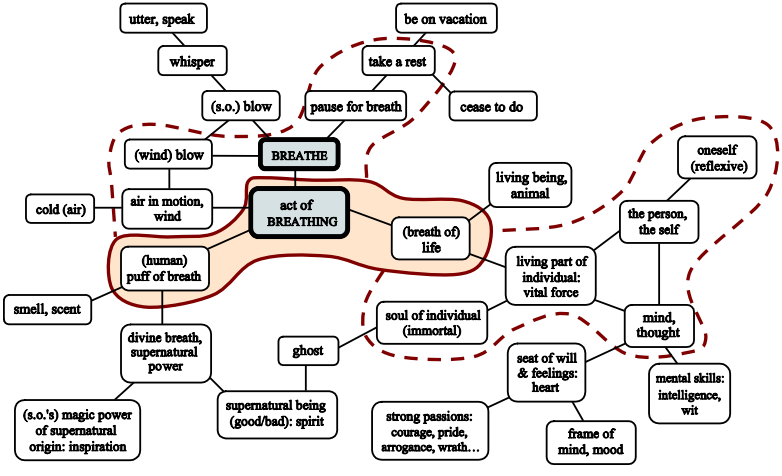


Figure 16 – Isolectic set for Standard Arabic *nafas*

Table 2 – Lexical data on the polysemy of {BREATHE}

	SANSKRIT	GREEK	GREEK	LATIN	LATIN	RUSSIAN	MANDARIN	ALEUT+	NAHUATL	MWOTLAP	NELEMWA	ARABIC	ARABIC	BEDJA	SAR
	ātman	*psūkhē*	*pneuma*	*anima*	*spīritus*	*dux*	*qì*	*anri-*	*imi'iyo*	*mōkhe-*	*horêâ-*	*rūḥ*	*nafas*	*šūk*	*koo*
BREATHE	[+]	[+]	[+]		[+]	[+]		[+]		[+]	+		[+]		+
(s.o.) blow		[+]	[+]		[+]	[+]		[+]		[+]	+		[+]		+
whisper, utter								+			+				
take a rest						[+]			+	[+]	+	[+]	[+]		[+]
be on vacation						[+]				[+]					
cease to do											+				
(wind) blow			[+]		[+]	[+]						[+]			
air, wind	[+]	[+]	+	+	+	[+]	+					[+]			+
cold (air)		[+]													
puff of breath	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
smell, scent			+		+	[+]	+		+	+		[+]			
ACT OF BREATHING	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
(breath of) life	+	+	+	+	+	+	+	+		+	+	+			
living being, animal	[+]	[+]		[+]		[+]									
vital force of individual	+	+	+	+	+	+	+	+		+		+	+	+	
person; self	+	+		+	+							[+]	[+]		
oneself (reflexive)	+											[+]	[+]		
mind, thought	+	+	+	[+]	+	+	+	+				+	[+]		
intelligence, wit	+	+													
will and feelings: heart		+	+	[+]	+	+	+					+			
pride, arrogance, wrath			+	[+]	+	+	[+]								
frame of mind, mood		+		[+]	+	+	+					+			
soul of indiv. (immortal)	+	+		+	+	[+]		+				+		+	
ghost		+		+		+		+				+			
divine breath or power			+		+	+						+			
magic power, inspiration			+		+	[+]						+			
supernatural being, God	+		+		+	+						+			

Figure 5 – The semantic map of {BREATHE} [Winword format]

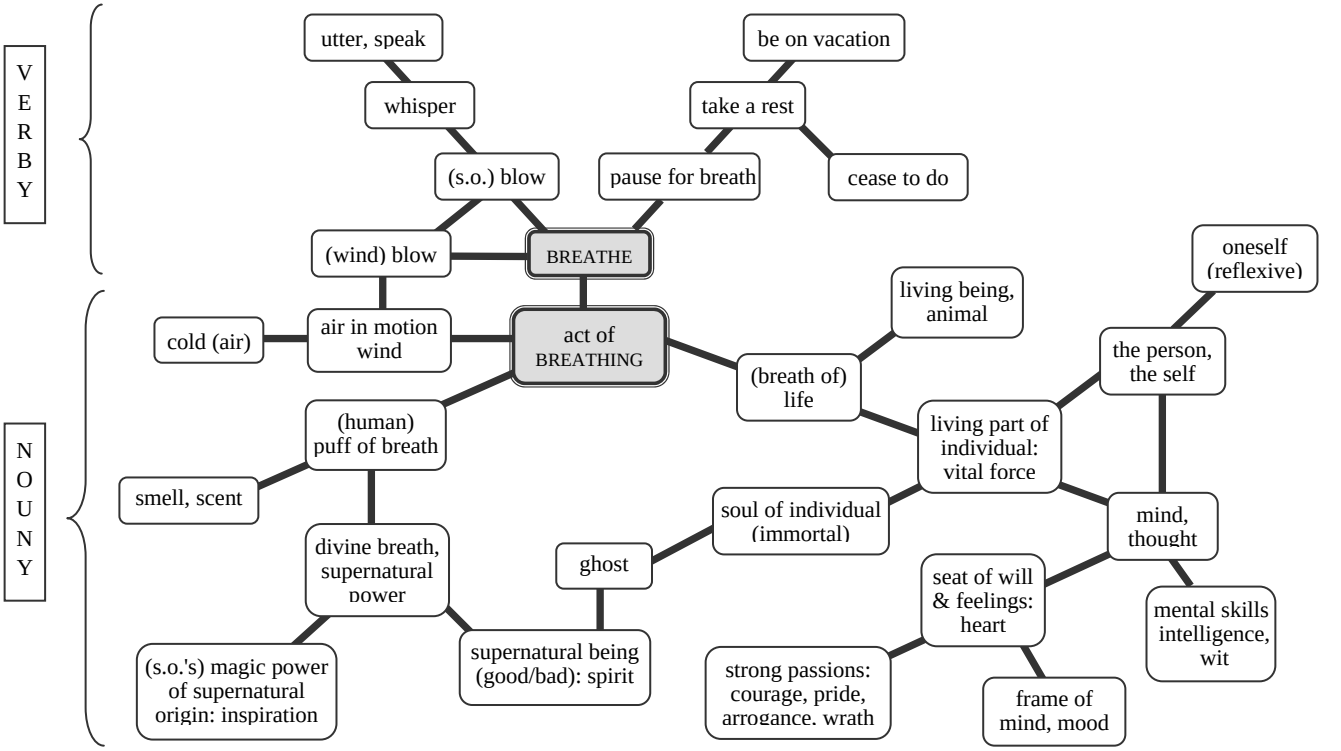
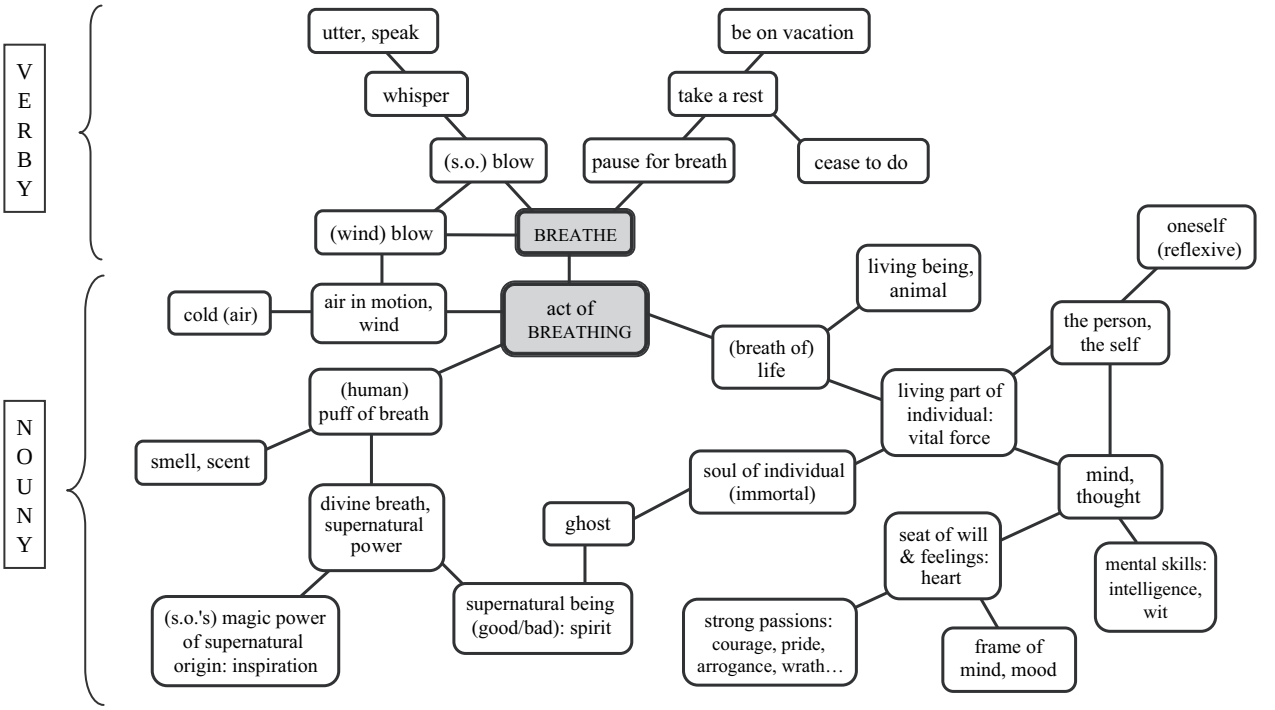



Figure 5 – The semantic map of {BREATHE} [EPS format]