Regular homophones: a tool for semantic typology and for linguistic reconstruction

Konstantin Pozdniakov, Guillaume Segerer

Konstantin Pozdniakov & Guillaume Segerer

Regular homophones:
a tool for semantic typology and for linguistic reconstruction

1 Introduction

This paper describes the use of a large lexical database to explore the phenomenon of regular homophony, as well as its value for linguistic reconstruction, based on a large sample of African languages. For a better understanding of our purposes, a few definitions are in order.

1.1 Definitions

- We use the term 'polysemy' when a single form has different, related meanings, for example: Bambara *kólǫ́* 'naked' ; 'empty'.
- We use the term 'homonymy' when a single form has different, unrelated meanings, for example, Abanyom *-tə̂ŋ* 'ashes' ; *-tə̂ŋ* 'ear'. Of course, the relationship between different meanings may not be known and there will be many cases when it is not easy to decide whether a form is polysemic or homonymic.
- We use the term 'homophony' to encompass polysemy and homonymy, i.e. when a single form has different meanings, independently of them being related. Thus, Bambara *kólǫ́* and Abanyom *-tə̂ŋ* are both homophones, the former being an instance of polysemy, the latter being an instance of homonymy.

These definitions are not intended to reflect any theoretical orientation about semantics or lexicology. Instead, they are simple tools that will help us to process large quantities of data extracted from a large lexical database. Our use of the term 'homophony', although not standard,[1] is preferred for its iconic force to the recently coined 'colexification' (François 2008) which in our opinion tends to lean too much towards polysemy.

1.2 Using regular homophones for semantic typology and semantic reconstruction

It is well known that the comparative method is very powerful in reconstructing forms, but not so when it comes to meanings. Unlike phonetic reconstruction – which is based on regular and systematic correspondences – semantic reconstruction relies on typological parallels only. These parallels are not only used in semantic reconstruction, but are also employed to justify cognate sets that include words with different meanings. As Emile Benveniste (1954: 251) puts it,

> "…en matière de sens, on n'a pour guide qu'une certaine vraisemblance, fondée sur le « bon sens », sur l'appréciation personnelle du linguiste, sur les parallèles qu'il peut citer. Le problème est toujours, à tous les niveaux de l'analyse, à l'intérieur d'une même langue ou

[1] It is actually difficult to assess the standard meaning of 'homophony'. It seems to be usually understood to be a subtype of homonymy, namely that which does not assume identical spelling. Thus, it is *de facto* reserved to languages with an orthography: "homophones: forms which are homonyms, at least phonetically. Often therefore of ones spelled differently: e.g. tier and tear [tɪə]." (Matthews 2007: 178). There is no entry for 'homophony' in the 5664-page long Encyclopedia of Language and Linguistics (Asher Ed., 1994), but under 'homonymy' one can read : "[bank 'slope, elevation in sea or river bed' and bank 'financial institution'] are at the same time homophones (distinct words with the same pronunciation) and homographs (distinct words with the same spelling)", although later in the same paragraph it is stated that: "The terminological relationship between homonymy, homography, and homophony is not entirely clear".

aux différentes étapes d'une reconstruction comparative, de déterminer si et comment deux morphèmes formellement identiques ou comparables peuvent être identifiés par leur sens"[2].

But the use of 'common sense' is often misleading. For this reason, the last decades have seen various attempts at creating typological databases dedicated to semantic shifts – whether attested or assumed – in the world's languages (Zalizniak *et al* 2012, Gévaudan *et al* 2003 among others). Although these projects have lead to interesting results, so far, no database of semantic shifts has been made available. Anna Zalizniak (2008: 219) notes that:

> "The idea of a collection of semantic shifts which are reiterated in different languages is not absolutely new. In 1964 a Russian etymologist Oleg Trubačev put forward the idea of a "Semasiological dictionary of Indo-European languages" with historically witnessed semantic shifts as lexicographic entries (Trubačev 1964). But this idea was never realized. A similar project was put forward by a German linguist Johannes Schröpfer in 1952 (but not actually realized, cf. Schröpfer 1979). Some other works have to be mentioned here, namely: Javorskaja (1992), Gak (1998), Tolstaja (2002), Sakhno (1998, 1999, 2001), Hénault-Sakhno & Sakhno (2001, 2005), Heine & Kuteva (2002), Koch (2003, 2004)".

Why is the creation of such a database so complicated? Let us examine the kinds of information that it is supposed to contain. According to A. Zalizniak *et al* (2012), lexical semantic shifts have the following possible realizations:

1) Synchronic polysemy: French *appeler* 'to call, to name'; Spanish *esperar* 'to wait, to hope'
2) Diachronic semantic evolution: Latin *caput* 'head' > French *chef* 'chief'
3) Morphological derivation: German *zählen* 'to count' > *erzählen* 'to tell'
4) Cognates: Russian *mesto* 'place' ~ Polish *miasto* 'town'
5) Borrowing: Russian *sortir* 'toilets' < French *sortir* 'to go out'

Cases (2), (4) and (5) imply that the languages where these lexical semantic shifts occur are well known, a situation which is not that frequent. Therefore, the use of these cases is problematic in diachronic studies, where typological parallels are needed in order to justify cognate sets and reconstructions. In the absence of such parallels, these justifications can only rely on hypotheses, which have to be justified themselves.

In case (3), one needs a good knowledge of the synchronic morphology of a language. For Russian, there is an excellent morpheme dictionary (Kuznetsova 1986) where all the lexical roots are listed with all their derived forms. But this kind of dictionary does not exist for French, for example. In the case of lesser known languages of course, such a luxury is never available.

We therefore decided to rely on case (1), the easiest to tackle. Moreover, some semantic relationships that belong to cases (2)-(5) in some languages may well be attested as examples of case (1) in others. For instance French *demeurer* 'to dwell' comes from Latin *demoror* 'to stay too long, to linger' (2), but the original meaning is not lost entirely in French where *demeurer* also means 'to stay'.

This paper aims at presenting a new approach to two usually separate fields, namely phonological reconstruction and semantic typology. This approach allows us to:

(1) delineate the semantic fields of words in different languages in order to feed a general typology of semantic change;

(2) exploit these data for the needs of comparative historical linguistics;

[2] "...when it comes to meaning, the only guide is a certain likelihood based on 'common sense', on the linguist's personal judgment, on the parallels they can provide. At every level of analysis, the problem is always – whether in a single language or at various stages of comparative reconstruction – to determine whether and how two formally identical or similar morphemes may be identified on the basis of their meaning."

Section 2 presents our methodology. In Section 3, we present preliminary results obtained by our method. We will deal not only with polysemic links in the languages of Africa, but also with the methodological aspects of our approach and the limits of its application. In the appendix (§7), we will give an inventory of the polysemic links for notions included in the Swadesh-100 list.

1.3 Using regular homophones to find of sound correspondences

Section 4 is devoted to the application of our approach to one of the main goals of comparative historical linguistics, namely the finding of regular sound correspondences. At the beginning of our research, we did not expect to use homophony as a tool for phonological reconstruction. It turns out, as we will show, that the analysis of regular homonyms offers unique opportunities for historical phonology.

In Section 5, we discuss the consequences of these findings for a better understanding of semantic shift, phonological reconstruction and the study of borrowings.

2 Methodology

Our approach is based on a concept that we call 'regular homophony'. Regular homophony is attested when two or more languages show an identical case of homophony (same form, different meanings). Thus, it includes both the cases of regular polysemy and the cases of regular homonymy. This allows for automatic searches in a large lexical database, which is the first step of our approach. The next step consists in separating the cases of polysemy from the cases of homonymy, using criteria that are discussed below (section 5.2).

The database used for this research is RefLex[3], the world's biggest multilingual online lexical database, which includes 1,150,000 words from 1289 sources covering 785 African languages (as of May, 2018). Further, G. Segerer wrote a program that extracts all homophones for a given meaning and sorts them by decreasing frequency. This program is not restrictive, i.e., besides true homonymic and polysemic forms, it retrieves items that may be considered as "noise" (cf. § 5.2). Therefore, results are sorted manually. The program selects two different kinds of homophones: i) items that are given different meanings in a single dictionary entry, and ii) identical items that represent distinct dictionary entries. Furthermore, the default output of the program lists 'regular' homophones only, that is, those that are represented in two or more sources. It will be argued that such homophones, even if attested in two languages only, are very unlikely to occur by chance. Unique cases (homophones attested in one source only) may still be retrieved by changing the program parameters.

The table below shows the 'wait' = 'hope' parallel[4] as found in the RefLex database (also a well-known case of polysemy in several Romance languages, as cited above):

LANGUAGE - Phylum[5]:Family	Form	Meanings
Manjaku - NC:Atlantic	*aʈaran*	'wait confidently', 'hope'
Joola Kasa - NC:Atlantic	*kob*	'wait', 'hope'
Kabiye - NC:Gur	*ɖáŋɔ̀ɔ̀*	'wait', 'hope', 'prevent'
Moyobe - NC:Gur	*pì-mɛ́*	'wait', 'hope'
Sura - AA:Chadic	*cān cīn*	'wait', 'hope', 'plan', 'want'

Table 1: Regular polysemy 'wait' = 'hope'

This parallel is attested in 6 different languages from 4 families in two phyla.

[3] Segerer & Flavier 2011-2018. www.reflex.cnrs.fr/database.

[4] For the need of the present paper, all the translations are given in English, but the sources in the database are mostly in French. This does not affect the overall content of the paper.

[5] Language phyla are abbreviated as follows: AA: Afroasiatic; NC: Niger-Congo, NS: Nilo-Saharan.

The program plots all semantic links for a given concept. Thus, the concept 'empty' yields the following 'regular homophony' (2 sources or more, the number of sources being given in brackets):

'naked' (14)	'to dry up' (4)	'simply' (2)
'dry' (11)	'vain' (4)	'null' (2)
'nothing' (9)	'hole' (4)	'clean' (2)
'useless' (8)	'be clean' (4)	'free' (2)
'simple' (8)	'open' (3)	'in vain' (2)
'desert' (5)	'empty-handed' (3)	'free (time)' (2)
'hollow' (4)	'not important' (2)	'innocent' (2)
'pure' (4)	'void' (2)	'sane' (2)

Table 2: Regular polysemy of 'empty'

Obviously, some of these items may be grouped together and reinforce each other, as for instance 'vain, in vain, not important, useless'. In fact, such groupings do appear (partially or totally) in the original translations in the database, as seen in Table 3 below. The relationship between 'empty' and 'vain' is seen as a semantic shift ('empty' > 'vain') in Zalizniak (2008: 228), with illustrations in Russian, Latin and Greek. In the RefLex database this polysemy can be automatically retrieved in 11 languages from various stocks, as shown in Table 3:

Language	Form	Meanings
Bedja - AA:Cushitic	*míʃuub*	'**vain**', '**empty**', '**useless**', 'idle', 'sarcastic (esp. talk)'
Hamar - AA:Omotic	*gúuri*	'**empty**', '**useless**', 'emptyhanded'
Zodi - AA:Chadic	*fòpláŋìì*	'**void**', '**useless**'
Joola Fogny - NC:Atlantic	*tikor*	'arid', 'naked', 'pure', '**vain**', '**empty**'
Ben Tey - NC:Dogon	*kòrŏy*	'simply', 'for nothing', '**in vain**', '**uselessly**', '**empty**'
Yorno-So - NC:Dogon	*kóóló*	'trivial, frivolous (e.g. talk) ', '**empty** (container) ', 'deserted (village) ', '**useless**, good-for-nothing (person)'
Yemba - NC:Benue-	*ndém*	'**useless**', 'nothing', '**empty**'
Gikyode - NC:Kwa	*gìyã̄ŋ*	'**empty**', '**in vain**'
Gola - NC:isolate	*fóólo*	'**empty**', '**vain**', 'for free', 'bowl'
Mano - NC:Mande	*fēlē*	'**uselessly**', 'simply', '**empty**', '**vain**'
Mundu - NC:Ubangi	*kōkōrō*	'**empty**', '**vain**'
Sereer - NC:Atlantic	*balal*	'**empty**', '**useless**'

Table 3: detailed polysemy of 'empty' = '(in) vain' = 'useless(ly)'

Table 3 represents a minimum, for reasons pertaining to the level of accuracy of the sources and to the way meanings are encoded in the database. A refined search in RefLex gives the following additional results:

Language	Form	Meaning
Lendu - NS: Central Sudanic	*pá*	'free', 'available', '**empty**', '**in vain**'
Gevia - NC:Bantu B	*kambure*	'**empty**', 'nothing', '**vain**'
Ndebele - NC:Bantu	*ze*	'**empty**', '**vain**', 'naked', 'void'
Dholuo - NS: Nilotic	*nɔ̀ɔ̀nɔ́*	'**useless**', 'empty-handed', '**empty**'
Nembe - NC:Ijoid	*pàpàbìó*	'**empty**', 'containing nothing', 'devoid of furniture or inmates', 'without loads', '**vain**', 'addled', 'futile', 'fruitless'

Table 4: detailed polysemy of 'empty' = '(in) vain' = 'useless(ly)' (continued)

There are important differences between our results and the semantic shifts as illustrated in Russian, Latin and Greek by A. Zalizniak:

First, given the size of our database, it is no surprise that we can show many more attestations of this polysemy. We found it automatically in 10 linguistic families, and manually in 4 additional ones. This excludes a random homonymy.

Second, our results do not depend on etymological hypotheses, but come from synchronic facts.

Third, this polysemy ('empty' = 'vain' = 'useless') may be assessed within a multidimensional semantic context around the concept 'empty'. It can be noted, for instance, that 'empty' is consistantly linked with 'naked' (13 languages in 9 families, Table 5) and 'dry' (13 languages in 8 families, Table 6):

Language	Form	Meanings
Joola Fogny - NC:Atlantic: Bak	*tikor*	'arid', '**naked**', 'pure', 'vain', '**empty**'
Manjaku - NC:Atlantic: Bak	*jənts*	'**naked**', 'pure', '**empty**'
Mankanya - NC:Atlantic: Bak	*p-jinţ*	'**to be empty**', '**to be naked**'
Fula (Maasina) - NC:Atlantic: North	*ɓol*	'alone', '**naked**', 'void', '**empty**', 'pure', 'simple'
Temne - NC:Mel: North	*ţin*	'**naked**', '**empty**'
Kisi - NC:Mel: South	*fóndò*	'dry', '**empty**', '**naked**'
Mumuye - NC:Adamawa	*wāātī*	'dry', '**empty**', '**naked**'
Nomaande - NC:Benue-Congo	*ètémbí*	'**empty**', '**naked**'
Nugunu - NC:Bantu	*gɔ̀dɔ̀ɔ́*	'**empty**', '**naked**', dry'
Bwamu - NC:Gur	*'ūé ~ 'ūá*	'**empty**', '**naked**'
Tommo-So - NC:Dogon	*ísé*	'**naked**', '**empty**', free (time)'
Mandinka - NC:Mande	*kénsèŋ*	'**naked**', '**empty**'
Bambara - NC:Mande	*kólo̲*	'**naked**', '**empty**'
Soninke - NC:Mande	*dúurò*	'**naked**', '**empty**', without anything'
Zarma - Songhay	*kóónù*	'**empty**', '**naked**'

Table 5: detailed polysemy of 'empty' = 'naked'

Language	Form	Meanings
Joola Fogny - NC:Atlantic: Bak	*tikor*	'**arid**', 'naked', 'pure', 'vain', '**empty**'
Balant - NC:Atlantic: Bak	*sɔ́lɛ̀*	'**dry**', '**empty**'
Krobu - NC:Kwa	*ŋm̀kpré*	'**dry**', '**empty**'
Kenyang - NC:Benue-Congo	*éwɔ̀*	'**dry**', '**empty**'
Nugunu - NC:Bantu	*gɔ̀dɔ̀ɔ́*	'**empty**', 'naked', '**dry**'
Kisi - NC:Mel: South	*fóndò*	'**dry**', '**empty**', 'naked'
Krim - NC:Mel: South	*hwày*	'**dry**', '**empty**'
Mumuye - NC:Adamawa	*wāātī*	'**dry**', '**empty**', 'naked'
Geji - AA:Chadic	*kōōnî*	'**dry**', '**empty**', 'hollow'
Wolaytta - AA:Omotic	*mela*	'**dry**', '**empty**'
Gofa - AA:Omotic	*mela*	'**dry**', '**empty**'
Gamo - AA:Omotic	*mela*	'**dry**', '**empty**'
Dawro - AA:Omotic	*mela*	'**dry**', '**empty**'

Table 6 detailed polysemy of 'empty' = 'dry'

Less attested parallels are also interesting, like 'empty' = 'desert', 'empty' = 'hollow', 'empty' = 'pure', etc. These parallels, although less frequent, are very unlikely to be due to chance. As the tables above show, rare meanings (*hollow*, *pure*) are often attested along with common ones.

The distribution of meanings as shown by the automatic search is of great importance when assessing the directionality of the evolution of meaning. For instance, in the whole database, there are only three instances of the 'naked' = 'dry' polysemy, all of which also show the meaning *empty* (Kisi, Mumuye and Nugunu, see Table 6). Therefore, 'empty' may be considered as a pivot meaning linking 'naked' and 'dry'. This assumption is made possible by the very large number of languages in the database. Even if the results cannot be considered as exhaustive[6], semantic parallels that are not revealed by our method may be viewed as atypical for African languages.

Finally, our method retrieves unexpected parallels that were unlikely to be postulated and examined. Some examples will be given below.

3 Homophony coming from polysemy

3.1 The use of regular polysemy in historical linguistics

Let's look at the regular homophony for the meaning 'nose'. Here are the most frequent homophones: 'nasal mucus' (10), 'life' (9), 'nostril' (4), 'sprout' (4), 'forehead' (3), 'soul' (3), 'root' (2), 'tip' (2) and 'bone' (2). Now let's have a closer look at one of these parallels, namely 'nose' = 'life':

language	Form	Meanings
Wolof - NC:Atlantic: North	*bakkan*	'nose', 'life', 'life breath'
Pepel - NC:Atlantic: Bak	*yil*	'nose', 'to live'
Joola Kasa - NC:Atlantic: Bak	*è-ɲẽndu*	'nose', 'course of lifetime'
Tengu-kan - NC:Dogon	*kɪni*	'nose', 'life'
Tommo-So - NC:Dogon	*kínú*	'nose', 'life'
Sembla - NC:Mande	*bὺnὺ*	'nose', 'life'
San (Maya) - NC:Mande	*ɲinde*	'nose', 'life'

Table 7: Regular polysemy 'nose' = 'life'

The 7 languages in Table 7 belong to three remotely related families of the Niger-Congo phylum, and chance cannot be the reason for such a distribution. Therefore, there must be a conceptual link between *nose* and *life* that is responsible for the observed polysemy[7]. It follows to reason that, when searching for cognate forms for *nose*, to look as well for forms meaning *life* or *to live*. Let us now look at one of the cognate sets for the meaning *nose* in the Atlantic languages[8] of the Niger-Congo phylum:

Language	Form	phonetic alignment									
Fula (Adamawa)	kin-e					k	i	n	-e		
Fula (Masina)	hin-ere					h	i	n	-ere		
Fula (Futa Jalon)	kin-al					k	i	n		-al	
Jaad	ɲa-siŋ				ɲa-	s	i	n			
Biafada	ɲa-sin				ɲa-	s	i	n			
Gunyaamolo	ɲan-kɪnd				ɲan-	k	ɪ	nd			
Gubaher	gu-cind			gu-		c	i	nd			
Kobiana	gu-ɲi-kin			gu-	ɲi-	k	i	n			
Wolof	bakkan		ba			kk	a	n			
Basari	ɛ-cɛ́n	ɛ-				c	ɛ́	n			
Bedik	ɛ-cʌ̀l	ɛ-				c	ʌ̀	l			
Konyagi	ì-cə̀l̃	ì-				c	ə̀	l̃			

[6] Africa has more than 2,400 languages, the database has only 785, many of them being represented by short wordlists only.

[7] 'breath' could be such a link. But there is no attested polysemy of the kind 'nose / breath'.

[8] Atlantic is a family of *ca.* 40 languages spoken on the westernmost coast of Africa, from Senegal to Guinea.

Joola Kwaatay	e-hinu	e-				h	i	n			-u
RECONSTRUCTION	*?-KIN					*k	*i	*n			

Table 8: 'nose' in the Atlantic languages: phonetic alignment of forms and reconstruction

The reconstructed form evokes the forms found in Dogon, a very distantly related group of *ca.* 20 languages:

Language	Form	phonetic alignment						
Yorno-So	kín	k	í	n				
Tommo-So	kínú	k	í	n	u			
Tengu-Kan	kɩni	k	ɩ	n		i		
Togo-Kan	kír̃í	k	í	r̃		í		
Toro-Tegu	cìr̃ò-ká	c	ì	r̃	ò		-ká	
Ben-Tey	cír̃ì	c	í	r̃		ì		
Jamsay	cír̃é	c	í	r̃		é		
Perge-Tegu	kír̃é	k	í	r̃		é		
Nanga	kír̃ê	k	í	r̃		é		
Bankan-Tey	cír̃è	c	í	r̃		è		
Najamba	kìnjáà	k	ì	nj				-áà
Mombo	kínjà	k	í	nj				-à
Yanda-Dom	kìnzà	k	ì	nz				-à

Table 9: 'nose' in the Dogon langages: phonetic alignment of forms

With the help of the polysemy shown above, we may also add a word of Limba (an isolate within the Niger-Congo phylum): *kin* 'to live; to be'. The inclusion of Limba *kin* 'to live' in a cognate set for 'nose' is supported by the synchronic attestation in various NC languages of the polysemy 'nose' = 'life'. Thus, this probable NC root is now attested in three independant West African branches of the phylum (instead of two), which gives stronger support for its reconstruction.

3.2 The use of regular polysemy in semantic typology

Regular polysemy is also very powerful in the domain of semantic typology. Let's look at a few examples:

3.2.1 'eye' = 'seed'

The regular homophony for the notion 'eye' is as follows (in brackets, the number of sources where it is attested): 'face' (101), 'seed' (12) , 'front' (9), 'color' (5), 'jealousy' (3). The polysemy 'eye' = 'face' = 'front' is quite trivial and does not need to be illustrated. The second more frequent polysemy for 'eye', namely 'eye' = 'seed', seems less obvious to us. What we mean here is that if we consider as cognates a form meaning 'eye' in language 1 and a form meaning 'face' in language 2, not much justification is required. But if we consider as cognates a form meaning 'eye' in language 1 and a form meaning 'seed' in language 2, we feel that an explanation is needed, and the fact that this polysemy is widely attested constitutes such an explanation. This subjective difference in the approach of the two homophonic pairs 'eye' = 'face' and 'eye' = 'ground' is in fact objectively reflected in the numbers of occurrences of these pairs in our corpus. While 'eye' = 'face' is attested in 101 different sources, only 12 sources show the pair 'eye' = 'seed' (see Table 10 below). It can then be assumed that if the data are processed with caution (i.e. the sample of sources is balanced, or the number of families is considered instead of the number of sources), then the frequencies of homophonic pairs will adequately reflect the semantic closeness of the respective pairs of meanings. It is no surprise, then, that the most frequent cases are those of polysemic relationships, while the homonymic ones will probably be quite rare.

Language	Form	Meanings
Day - NC:Adamawa	*nɔ́n*	‘**eye**’, ‘face’, ‘front’, ‘**seed**’, ‘to tire’
Mumuye - NC:Adamawa	*nūn*	‘**eye**’, ‘face’, ‘capable’, ‘**seed**’
Gula Iro - NC:Adamawa	*ííl (è)*	‘**eye**’, ‘name’, ‘**seed**’
Baka - NC:Ubangian	*là*	‘to sleep’, ‘child’, ‘**eye**’, ‘who? ’, ‘**seed**’
Sango - NC:Ubangian	*le*	‘face’, ‘**seed**’, ‘**eye**’
Bokobaru - NC:Mande	*wɛ́ɛ*	‘**eye**’, ‘appearance’, ‘**seed**’, ‘demand’
Kele - NC:Bantu B	*diʃ*	‘**eye**’, ‘**seed**’
Shiwa - NC:Bantu A	*mi*	‘**eye**’, ‘**seed**’
Eloyi - NC:Benue-Congo:Idomoid	*kúyī*	‘**eye**’, ‘**seed**’
Ik - NS:Eastern Sudanic	*èkú*	‘**eye**’, ‘**seed**’, ‘headlight’
Hamar - AA:Omotic	*áapi*	‘**eye**’, ‘**seed**’, ‘fruit’
Basketo - NC:Omotic	*ááp*	‘**eye**’, ‘fruit’, ‘**seed**’

Table 10: ‘eye’ = ‘seed’

This polysemy is widely attested in Africa, since it is found in 6 mid-level linguistic families and 3 phyla.

As previously said, the automated search is not exhaustive, since 2/3 of African languages are absent from the database. Therefore, the polysemies found in RefLex can be expected to be found in other sources. Indeed, ‘eye’ = ‘seed’ is found in Fon (NC:Kwa) *nùkún* and in Degema (NC:Benue-Congo:Edoid), where ‘eye’ is *ɔ́ŋʷɔ́ ádʊ*, literally meaning ‘seed-face’, suggesting that the polysemy could be a case of semantic shift, from ‘seed’ to ‘eye’.

3.2.2 ‘eye’ = ‘colour’

Language	Form	Meanings
Dendi - Songhay	*mò*	‘eye’, ‘colour’
Dan (gwèèta) - NC:Mande	*yɡ̃*	‘eye’, ‘colour’
Mayaa - NC:Mande	*yɛrɛ*	‘eye’, ‘colour’
Obolo-Andoni - NC:Benue-Congo	*cyèéŋ*	‘eye’, ‘colour’
Bassa - NC:Kru	*jéɖé*	‘eye’, ‘colour’

Table 11: ‘eye’ = ‘colour’

This polysemy is far from being obvious. However, it is attested in at least 4 language families, which excludes chance. This example shows that there is often a ‘missing link’ that can help understand an otherwise unexpected polysemy. The systematic study of the polysemy around ‘eye’ and ‘colour’ suggests that this link originates in the chained polysemy ‘eye’ = ‘sight’ = ‘aspect’ = ‘species, type’ = ‘colour’.

Within RefLex, this chained polysemy is attested as a whole in Obolo-Andoni, and partially in a few other languages as shown in Table 12:

Language	Form	Meanings
Obolo-Andoni - NC:Benue-Congo	*cyèéŋ*	‘eye’, ‘view, sight’, ‘colour’, ‘type, genre’
Karekare - AA:Chadic	*idau*	‘eye’ , ‘face’
Gbaya-Yaayuwee - NC:Ubangi	*yí*	‘eye’, ‘face’, ‘aspect’
Ijo - NC?:Ijoid	*tɔ̀rɔ́*	‘eye’, ‘face’, ‘class’, ‘kind’, ‘type’…
Gevia - NC:Bantu B	*mo-nɔngɔ*	‘quality, aspect’, ‘species, sort’, ‘behaviour’, ‘colour’

Yorno-So - NC:Dogon	*sîì*	‘type, sort, kind’, ‘color’
Yanda Dom - NC:Dogon	*tɔ̀gù*	‘type, sort, kind’, ‘color’
Zaar - AA:Chadic	*kálā*	‘Colour’, ‘Type’

Table 12: chained polysemy from ‘eye’ to ‘colour’

Outside RefLex, one may find a very good illustration of the chain in Christaller’s (1933) voluminous Akan dictionary:

aní ‘the eye or eyes; a look’
‘the face, visage, countenance’
‘the face, front, forepart’
‘the face, surface’
‘the face, visible part’
‘colour’
‘opening for a fountain’

3.2.3 ‘tooth’ = ‘name’

Some cases of regular homophony are difficult to assign to either polysemy or homonymy. For instance, several African languages show a homophony between the meanings ‘tooth’ and ‘name’.

Let’s start with the Gbaya languages (a cluster of closely related languages in the Ubangi branch of the Niger-Congo phylum). A thousand Proto-Gbaya words have been reliably reconstructed by Moñino (1995). The homophony ‘tooth’ = ‘name’ is found in half of the languages in the cluster, as shown in Table 13:

Gbaya lect	‘tooth’	‘name’	Gbaya lect	‘tooth’	‘name’
Kara Ɓodoe	*ŋín*	*ŋín*	Bangando	*ŋîì*	*lí*
Kara Ɓokpan	*ŋín*	*ŋín*	Manza	*gòkò*	*l̰ī*
Ɓokoto	*ŋín*	*ŋín*	Ngbaka-Min.	*gòkò*	*lī*
Ɓozom	*ŋín*	*ŋín*	ʔali	*ŋīnī*	*lī*
Gbeya ƁoʔoroBofi	*ŋín*	*ŋín*	Ɓofi	*nīní*	*líŋ*
Gbanu	*ŋínì*	*ŋínì*	Ɓiyanda	*yínì*	*líŋ*
Kara Ɓugui	*nín*	*nín*	Ɓuli	*yínì*	*líŋ*
Yaayuwee	*nín*	*nín*	Toongo	*yínì*	*líŋ*
Kara Ɓoŋina	*nín*	*nín*	Lai	*nín*	*líŋ*
Ɓoya	*nín*	*nín*	Mbodomo	*ŋínì*	*líŋ*

Table 13: ‘tooth’ and ‘name’ in the Gbaya cluster

In the left column, the forms for ‘tooth’ and ‘name’ are exactly the same. In the right column however, the forms differ but remain similar (except for Manza). The reconstructions for ‘tooth’ and ‘name’ are **ŋín* and **l̰íŋ* respectively, implying that in a half of the Gbaya lects the forms have converged. Another scenario could be that the forms were homophones in Proto-Gbaya and diverged in some of the lects. This alternate scenario finds support from the forms of homophones with the same meanings in other Niger-Congo languages (Table 14):

Language	Form	Meaning
Kota - NC:Bantu B	*jinɔ*	‘tooth’, ‘name’
Bushong - NC:Bantu B	*dìín*	‘tooth’, ‘name’
Pindi - NC:Bantu C	*e-dzín*	‘tooth’, ‘name’
Samba Leko - NC:Adamawa	*nə́gə́l*	‘tooth’, ‘name’
Gbaya Ɓozom - NC:Ubangi	*ŋín*	‘tooth’, ‘name’
Attié - NC:Kwa	*hl̰̃*	‘tooth’, ‘name’
Nyabwa - NC:Kru	*ŋéné*	‘tooth’, ‘name’

Wobé - NC:Kru	*ɲnɛ̰́*	‘tooth’, ‘name’
Dogulu Dom - NC:Dogon	*ʔinu*	‘tooth’, ‘name’
Mombo - NC:Dogon	*ini*	‘tooth’, ‘name’

Table 14: ‘tooth’ = ‘name’ in Niger-Congo

Besides the Ubangi branch, the homophony ‘tooth’ = ‘name’ is attested in the Bantu, Adamawa, Kwa, Kru and Dogon branches. And as in Ubangi, a number of languages show for these meanings forms that are resembling each other without being exactly the same (Table 15):

Language	‘tooth’	‘name’
Ikaan - NC : Ukaan	*ò-ɲɪ̰̀*	*è-nɪ̰́*
Najamba - NC:Dogon	*ìnɔ̀ɔ́*	*ínèn*
Mombo[9] - NC:Dogon	*ínnì*	*íní*
Yanda Dom - NC:Dogon	*ìn*	*ín*
Koonzime - NC:Bantu A	*è-jyè*	*díè*
Sangu - NC:Bantu B	*dinu*	*dinə*
Likile - NC:Bantu C	*l-ínò*	*l-ínà*
Pana - NC:Gur	*yírí*	*yìré*
Eloyi - NC:Benue-Congo	*kú-ɲì*	*é-ɲí*
Adioukrou - NC:Kwa	*nɛ́n*	*nín*
Pere - NC:Adamawa	*núūlɛ̀*	*nīīlɛ̀*

Table 15: Close forms for the meanings ‘tooth’ and ‘name’ in Niger-Congo

This homophony, both complete and partial, is also attested in the South Bauchi languages, a cluster belonging to Chadic, a branch of Afroasiatic, i.e. not Niger-Congo (Table 16):

Language	‘tooth’	‘name’
Geji (Gezawa)	*sìn*	*sìn*
Dwot	*ʃín*	*ʃɪn*
Bu: (Zaranda)	*ʃɪn*	*ʃɪn*
Pelu	*ʃīn*	*ʃīn*
Zoɗi	*ʃèn*	*ʃɪn*
Geji 2	*ʃīŋ*	*ʃīn*

Table 16: ‘tooth’ = ‘name’ in Chadic

In some sources (Dwott, Zoɗi, Geji 2), the forms are non-identical, varying in tone and/or in segmental shape. As for Gbaya, two scenarios can be proposed to explain this: either the homophony can to be traced back to the Proto-language, and forms have diverged in some languages, of the forms were close but not identical in the Proto-language, and converged in some of the languages. What matters here is that these homophones occur both in Niger-Congo (several branches) and in Chadic, which can hardly be due to chance. Does it mean that this distribution suggests a kind of polysemy that is not yet understood? We will come back to this case after having examined the various interpretations of homophony.

3.3 Parallel polysemy

Our approach allows us to study pairs of homophones, but also clusters of polysemic notions. To illustrate this, let us look at words denoting cardinal points in African languages. The dominant model assimilates ‘North’ and ‘South’ to ‘up’ and ‘down’ respectively, and ‘East’ and ‘West’ to ‘right’ and ‘left’ respectively, as illustrated with Lega (NC, Bantu D):

[9] This is a different source to that cited in Table 14.

ko-ntara	‘North’	‘up’
ma-língá	‘South’	‘down’
i-dómɛ	‘East’	‘right’
i-kází	‘West’	‘left’

Table 17: Directions in Lega-Beya (NC:Bantu D)

The model may differ slightly, with ‘North’ being polysemic with ‘head’ or ‘mountain’ for example (Table 18):

Language	‘North’	‘South’
Wolof - NC:Atlantic	*bɔpp* ‘head’, ‘chief’	*taank* ‘foot, leg’
Mani - NC:Mel	*tòk* ‘high’, ‘on’	*làmbàn* ‘down, low’
Kwegu - NS:Eastern Sudanic	*gáàk* ‘mountain’	
Mano - NC:Mande	*lāā* ‘upstream’	*ɓēí* ‘downstream’
Nzadi - NC:Bantu B	*dùù* ‘sky’, ‘upstream’	*ntsé* ‘bottom’, ‘downstream’

Table 18: Extension of the model ‘North’ = ‘up’ and ‘South’ = ‘down’

Some languages show a different orientation model, where it is ‘East’ which is polysemic with ‘up’. ‘North’ is then logically associated to ‘left’, and so forth, as in Susu (Table 19):

kɔ̀ɔ̀lá	‘North’, ‘left’
yìrèfáɲì	‘South’, ‘right’
fúgé	‘East’, ‘up’, ‘highland’
láábé	‘West’, ‘down’, ‘lowland’

Table 19: The model ‘North’ = ‘left’ in Susu (NC:Mande)

This model is also partially attested in languages from other genetic stocks (Table 20):

Language	‘North’ = ‘left’	‘South’ = ‘right’	‘East’ = ‘up’	‘West’ =’down’
Nalu - NC:Atlantic	*kimaak*	*kilaŋ*		
Fula-Fuuta Jalon - NC:Atlantic	*nano*	*ɲaamo*		
Tangale - AA :Chadic	*kedè*	*sànùwa*		
Temne - NC:Mel	*mɛrɔ*	*diɔ*		
Ror - NC:Benue-Congo			*ūdɔ́m*	*ūtáʔ*
Limba - NC:isolate		*hotomi*		*kapoti*

Table 20: The model ‘North’ = ‘left’ in various languages

Other orientation models are rare, but they do occur. In Bangime, an isolate NC language, the model is West-oriented: North is on the right and South is on the left. Baga Mboteni has a comparable system (Table 21):

Language	‘right’ = ‘North’	‘left’ = ‘South’
Bangime - NC:isolate	*síí béè*	*báràà n nìì*
Baga Mboteni - NC:Atlantic	*i-rámì*	*i-rálómlà*

Table 21: West-oriented models

Even rarer are ‘horizontal’ models, where ‘North’ is not ‘up’ but ‘face’ or ‘front’. Such a system is attested in Obolo-Andoni (Table 22):

Form	Meaning	Origin
ágàŋ	‘side’, ‘direction’	

ágàŋ-ísí	'North', 'forward'	*ísí* 'face'
ágàŋ-ùdûŋ	'South', 'backward'	*ùdûŋ* 'back'
ágàŋ-ùlòm	'East', 'right'	*ùlòm* 'right'
ágàŋ-ùjìt	'West', 'left'	*ùjìt* 'left'

Table 22: A horizontal model: Obolo-Andoni (NC:Benue-Congo)

This kind of model may have a different 'reference' point. In Tupuri (NC:Adamawa), *pēlē* means 'front, face' and also 'West'. The polysemies 'right' = 'East' and 'left' = ''West' may co-occur with the more widespread polysemies 'right' = 'man' and 'left' = 'woman', as in Lega-Beya (NC:Bantu D): *ì-kází* 'left', 'woman', 'East' / *í-dómɛ̀* 'right', 'man', 'West'.

Finally, a few languages exhibit 'oppositive homophony', i.e., pairs of homophones with opposite meanings (Table 23):

Language	Form	Meanings
Fali Gili - AA:Chadic	*tàrkú*	'North' = 'South'
Hai - NC:Ubangi	*céká-ngàlân*	'North' = 'South'
Nomaande - NC:Benue-Congo	*èndáŋálɛ́ànà*	'East' = 'West' = 'width'

Table 23: Oppositive homophony

This type of homophony is quite frequent and is dealt with in more details in the next section.

3.4 Oppositive homophony

Oppositive homophony is attested in most languages families in Africa. Let us look at a few examples, beginning with the case of 'day' = 'night' (Table 24).

3.4.1 'day' = 'night'

Language	Form	Meanings
Koonzime - NC:Bantu A	*èsû*	'day' = 'night'
Wongo - NC :Bantu C	*butú*	'day' = 'night'
Bomboma - NC:Bantu C	*mòkɔ̀lɔ̀*	'day' = 'night'
Holoholo - NC:Bantu D	*sìgí*	'day' = 'night'
Nyambo - NC:Bantu J	*etʃiro*	'day' = 'night'
Nyankore - NC:Bantu J	*ekiro*	'day' = 'night'
Haya - NC:Bantu J	*ekílo*	'day (24 hours)' = 'night'
Kete - NC:Bantu L	*-tsùk*	'day' = 'night'
Ngbaka - NC:Ubangi	*tṵ̀*	'day' = 'night' = 'darkness'
Day - NC:Adamawa	*dùù*	'day' = 'night'
Mangbetu - NS:Central Sudanic	*nékìnì*	'day' = 'night'
Melokwo - AA:Chadic	*làvāŋ*	'day (24 hours)' = 'night'
Sura - AA:Chadic	*pār*	'day' = 'night'
Gyannzi (Geji) - AA:Chadic	*gās*	'day of the week' = 'night'

Table 24: Oppositive homophony 'day' = 'night'

This kind of homophony is difficult to analyse. Some dictionaries (see Haya and Melokwo above) add '(24 hours)' as a precision, thus suggesting that the meaning could be something like 'what separates a point in day from the same point one day before/after', which could apply to both 'day of 24h' and night. But most sources don not provide such a precision. What matters to the comparativist is that this regular polysemy allows them to compare the word for 'day' in a language (for example Bende *nsiku*) with a word for 'night' in another (for example Tswana *bʊ̀-síχʊ́*).

3.4.2 yesterday = tomorrow

This polysemy is attested in at least 53 languages of Africa, making it the most frequently attested oppositive homophony. In Proto-Bantu alone, five different roots showing this homophony are reconstructed (Table 25):

Form	Meanings
bàdì	'the next day/ the before day', '**yesterday**', '**tomorrow**'
dó	'**yesterday**', '**tomorrow**'
dóbí	'the next day/ the before day', '**yesterday**', '**tomorrow**'
kéɪda ~ kéɪdò	'morning', '**yesterday**', '**tomorrow**'
jàná	'**yesterday**', '**tomorrow**', 'day before yesterday', 'day after tomorrow'

Table 25: 'yesterday' = 'tomorrow' in Proto-Bantu (NC).

Besides Bantu, this polysemy is widely attested in various branches of the Niger-Congo phylum (Table 26):

Language	Form	Meanings
Yemba - NC:Benue-Congo	*èzó*	'**yesterday**', '**tomorrow**'
Alege - NC:Benue-Congo	*óófuue ~ kóófuue*	'**yesterday**', '**tomorrow**'
Ibibio - NC:Benue-Congo	*ǹkpɔ́ŋ*	'**yesterday**', '**tomorrow**'
Gbagyi - NC:Benue-Congo	*ósu*	'**yesterday**', '**tomorrow**'
Abure - NC:Kwa	*èhími̱*	'**yesterday**', '**tomorrow**'
Adioukrou - NC:Kwa	*éfì*	'**yesterday**', '**tomorrow**'
Alladian - NC:Kwa	*ɛ̄nɛ́*	'**yesterday**', '**tomorrow**'
Ebrie - NC:Kwa	*ɛ̀pì*	'**yesterday**', '**tomorrow**'
Eotile - NC:Kwa	*àsʊ́ɔ̀*	'**yesterday**', '**tomorrow**'
Dangme - NC:Kwa	*ɛzɔɔ*	'**yesterday**', '**tomorrow**'
Aja-Gbe - NC:Kwa	*eetsɔ*	'**yesterday**', '**tomorrow**'
Xwela-Gbe - NC:Kwa	*sɔ*	'**yesterday**', '**tomorrow**'
Godie - NC:Kru	*zɤ̄kū*	'**yesterday**', '**tomorrow**'
Nyabwa - NC:Kru	*glāápē*	'**yesterday**', '**tomorrow**'

Table 26: 'yesterday' = 'tomorrow' in Niger-Congo

Some languages may develop this polysemy from the meaning 'day', others will do it from the meaning 'night', or these meanings may arise from 'morning' or 'evening'. But the question remains: how can these two opposite meanings combine into a single form? In addition, many languages that show the 'yesterday' = 'tomorrow' polysemy also show the polysemy 'day before yesterday' = 'day after tomorrow' (Table 27):

Language	'yesterday' ='tomorrow'	'day before yesterday' ='day after tomorrow'
Ikizu - NC:Bantu E	*izo*	*ɓasirʸa*
Bungu - NC:Bantu F	*iyolo*	*isikʷinji*
Fipa - NC:Bantu F	*mutondo*	*uwanda unji*
Rungwa - NC:Bantu F	*iyolo*	*ilʸanda lindi*
Nyamwezi - NC:Bantu F	*igolo*	*mazʊʊli*
Mambwe - NC:Bantu F	*mutondo*	*mutondo ulʸa*
Ha - NC:Bantu J	*ejo*	*hidʸa yʼejo*

Gbagyi - NC:Benue-Congo	*ósu*	*ɥīgʸékʷó*

Table 27: 'yesterday' = 'tomorrow' and 'day before yesterday' = 'day after tomorrow'

It is worth noting that no language has been found with the 'yesterday' = 'day before yesterday' polysemy. All but one of the languages in Table 27 above belong to the Bantu stock. This is a good example of a regular polysemy with limited and local distribution, represented by forms that may not be cognates with each other.

Oppositive homophony is also (and maybe mostly) attested with verbs. Examples include 'to sell' = 'to buy', 'to take' = 'to give', 'to begin' = 'to stop', etc. The very fact that opposite meanings are rendered by the same forms suggests that it is a particular kind of polysemy, rather than fortuitous homonymy. In that respect, even unique cases can be said to show a certain type of regularity. This phenomenon is of course attested in other languages in other parts of the world (see French *louer* 'to rent' = 'to let').

Before proceeding with the second part of this paper, we wish to emphasize two points:

1) All the examples above show "selected" cases of homophony, i.e. those that are retrieved by the program, plus a few hand-picked ones. Many sources are to small to give the precision that we would wish for. For example, the cardinal points discussed above are far from being systematically present in the sources ('North' occurs in 126 sources out of 788).

2) It is often the case that different polysemies reinforce each other. For instance, 'seed' = 'species' is directly attested in only three languages (Table 28):

Language	Form	Meanings
Yemba - NC:Benue-Congo	*nzʷɛ̀t*	'**seed**', '**species**'
Bambara - NC:Mande	*sí*	'**seed**', 'generation', 'race, **species**, kind'
Dii - NC:Adamawa	*g̰ǵ̰*	'**seed**', '**species**', 'parents'

Table 28: 'seed' = 'species': direct homophony

But if we consider closely related meanings ('seed' > 'grain', 'species' > 'sort', 'kind', 'breed'), the cases of homophony are much more frequent (Table 29):

Language	Form	Meanings
Ma'di - NS:Central Sudanic	*ɔ̀rī*	'**seed**', 'crop', '**breed**'
Miya - AA:Chadic	*varay*	'**seed**', '**kind**, **type**', 'one's blood relatives'
Tiba - NC:Adamawa	*bì*	'**seed**', '**kind**'
Bangime - NC:Isolate	*búrúù*	'**seed**', '**kind**'
Akposo - NC:Kwa	*úkúlú*	'**seed**', '**sort of**'
Hausa (Ader) - AA:Chadic	*írìì*	'**seed**', '**genus**', '**type**'
Samo-Maya - NC:Mande	*dugo*	'**type**', '**seed**', 'parent'…

Table 29: 'seed' = 'species': indirect homophony

Likewise, the homophony 'seed' = 'parent' is directly found in only 2 languages (Table 30):

Language	Form	Meanings
Samo-Maya - NC:Mande	*dugo*	'type', '**seed**', '**parent**'…
Dii - NC:Adamawa	*g̰ǵ̰*	'**seed**', 'species', '**parents**'

Table 30: 'seed' = 'parent': direct homophony

But it can be justified by comparable parallels (Table 31):

Language	Form	Meanings
Miya - AA:Chadic	*varay*	'**seed**', 'kind, type', '**one's blood relatives**'
Mbuko - AA:Chadic	*zàhàv*	'**seed**', '**clan**', 'people'

Zarma - Songhay	*dúmí*	'**seed**', '**race**'
Mundu - NC:Ubangi	*ngùwà*	'**seed**', '**clan**', '**race**', 'neighbour'
Hausa (Ader) - AA:Chadic	*írìì*	'**seed**', 'genus', 'type', '**lineage**'
Samo-Maya - NC:Mande	*dugo*	'**type**', '**seed**', 'parent'…

Table 31: 'seed' = 'parent': indirect homophony

4 Regular homonymy as a tool for identifying regular sound correspondences

Polysemy is unable to explain all the cases of regular homophony. Many cases must be considered as instances of homonymy. At first, we excluded these cases from the scope of our research, but we finally decided to keep them apart, because we found they could be of great help fulfilling the most important needs of comparative historical linguistics. In this section, we intend to show how regular synchronic homonyms allow us to find regular sound correspondences between related languages. Moreover, we will show that the correspondences found with this method are of better quality than the ones given by "traditional" means. But let us take a few examples.

In the Ring cluster (belonging to the Grassfields group of the Benue-Congo branch of the Niger-Congo phylum), there are a few homophonous pairs for the following meanings: 'bee' = 'sun', 'sun' = 'to hear', 'to hear' = 'snake' and 'bee' = 'to skin'. Note that none of these homophonous pairs are attested outside the Ring cluster, and none of the pairs may safely be viewed as polysemic. Let us now look at the forms (words included in homophonous pairs are in grey cells)(Table 32):

Language	'bee'	'sun'	'to hear'	'snake'	'to skin'	
Weh	*zōw*		*ízʌ́w*	*zō*	*ízʌ̂w*	**z**
Aghem	*ízú*	*ízú*	*ízú*	*zúɣó*	*ízû*	**z**
Isu	*zó / zú*	*tsʊ̀ / tsɬ*	*zúʔ*	*zʊ̄ə̄ / zuā*	*zú*	**z**
Bafmeng	*ʌ̄zɬ́*	*īʒṵ́*	*zɬ*	*ʌ̄ʒó*	*zɬ̀*	**zl/ʒU**[10]
Babanki	*ʒṳ̀*	*ə̀cṵ́ ʒṵ́*	*ʒṵ́*	*ʒù*	*ʒṳ̌*	**ʒ**
Kom	*ə̄ʒvʊ́*	*īcvʊ́ īʒṵ̄*	*ʒvʊ̄*		*ʒvʊ̀*	**ʒv**
Oku	*ɥúɔ*	*cuî*	*ɥuɔ̂*	*ɥuɔ̂*	*ɥuɔ̀*	**ɥ**
Mbizinaku	*ə̄ɥú*	*ə̄cvʊ̂*	*ɥû*	*ə̄ɥuʌ̂*	*yù*	**ɥ**
Bum	*āɥū*	*ācwî*	*ɥû*	*āɥuɔ̂*	*yù*	**ɥ**

Table 32: Cross homophonies in the Ring cluster

We see here that there are different sequences of homophones in every language. But we also notice that for every language, the initial consonants are the same (or nearly the same) for each one of the 5 meanings. These are regular sound correspondences between forms that were phonetically close in proto-Ring, to the point that they sporadically converged to become homophones. We are not specialists of the Ring languages and we do not know which initial consonant should be reconstructed, but we have enough evidence to assert that it was the same consonant for each series and to set the regular sound correspondences between these languages[11].

[10]From the data in the table, one may assume that in Bamfeng, **z-** is observed before unrounded vowels, and ʒ- before rounded vowels.

[11] Actually, in the thesis from which the Ring data come from (Paulin 1995), this consonant is reconstructed as **d*.

Since these words existed in proto-Ring, and since the Ring languages are part of the Benue-Congo branch of NC, then we can expect to find correspondences with another branch of Benue-Congo, namely Bantu. In Table 33 we see the corresponding forms found in Proto-Bantu:

gloss	'bee'	'sun'	'to hear'	'snake'	'to skin'	*C
Proto-Bantu	*jíkì*	*jʊ́bà*	*jígu ~ júgʊ*	*jóká*	*jʊ̀b ~ jub*	***j**

Table 33: proto-Bantu reconstruction correponding to the Ring series of Table 32

All these words have the same initial consonant **j-* in proto-Bantu (a consonant which is absent from the forms in Ring). We can say that the study of (not so) regular homophony in Ring allowed us to find regular sound correspondences between two branches of Benue-Congo. From this point we can extend the search to other series where there is no homophony. Let us take the proto-Bantu words with an initial **j-* and search for the corresponding forms in Ring. This yields a few more series (Table 34):

	to sing	name	be dry	to buy	yesterday
Weh	*ízʊ́m*	*ízʌ́t*	*ízóm*	*ízó*	*ázɔ́ɔ́i*
Aghem	*ízɔ́m*	*ízín*	*ízúm*	*ízʊ́*	*ázɔɔ́*
Isu	*zém*	*ízét*	*zúm*	*zō*	*ázúā*
Bafmeng	*ʒɔ̄m*	*īzí*	*zōm*	*ʒɔīŋ*	*ʌ́zɔīŋɣê*
Babanki	*ʒə́m*	*ə̀ʒí?*	*zə́m*	*zén*	*zɔ́n*
Kom	*ʒɛ̄m*	*īɥín*	*yōm*	*ɥüīn*	*īɥɔ̄ní*
Oku	*ɥîm*		*ɥôm*	*ɥôn*	*īɥɔ̄nín*
Mbizinaku	*ɥə̂m*		*ɥôm*	*ɥuîn*	*īɥuʌ̄nə́*
Bum	*ɥʌ̂m*	*īɥə́t*	*ɥûm*	*ɥʊ̂n*	*íɥuɔ̄nī*
Proto-	***jímb***	***jínà***	***jʊ́m***	***jʊ̀d***	***jó, jànə́***

	to press	to yawn	to sweep	to kill	C1
Weh	*ízɔ̂?*	*ízaî*	*ízʌ̂t*		***z***
Aghem	*ízɔ̂*	*íɥə̂sɔ̀*	*ízê*	*íɥí*	***z, ɥ***
Isu		*zá*	*zʌ̀t*	*ɥíh*	***z, ɥ?***
Bafmeng	*zə̌?tə̂*	*ʒəfʌ̂*	*zʌ̀s*	*ʒṻ*	***z, ʒ***
Babanki	*zə́?tə́*	*zísə́*	*ʒèsə́*	*ʒwí*	***ʒ, z***
Kom	*yɛ̄?tə́*	*yàsì*	*ɥès*	*ʒṻ*	***ʒ, ɥ, y***
Oku	*ɥīktê*	*éɥaɥásé*	*ɥès*	*ɥuî*	***ɥ***
Mbizinaku	*ʒīktə̂*	*ɥuàsə̀*	*yètə̀*	*ʒvʊ̂*	***ɥ, ʒ, y?***
Bum	*ɥəktî*		*yʌ̀h*		***ɥ, y?***
Proto-Bantu	***jégɪk***	***jájì ~ jájò***	***jèjò 'broom'***	***jìp, jít***	***j***

Table 34: Ring/proto-Bantu sound correspondences

All the series do not show a perfect regularity, and there are still problems to solve. But we can nevertheless safely claim that: a) in the proto language above Bantu and Ring, all these forms had the same initial consonant; b) the homophony developped independantly in every Ring language and c) we have found sound correspondences between Ring and proto-Bantu without relying on proto-Ring.

Let us now turn to another case of "local regular homophony": 'new' = 'fire; be burnt'. This homophony is only attested in Bantu and in one non-Bantu Benue-Congo language. This case is interesting because the meanings involved are part of the so-called "basic lexicon" as exemplified by the famous Swadesh list. In Bantu we only find homophones like 'new' = 'ripe' and 'fire; be burnt' = 'ripe', but there are very few examples of the triple homophony 'new' = 'fire; be burnt' = 'ripe' (Table 35):

Language	‘fire; be burnt’	‘ripe’	‘new’
Ngwo - NC:Benue-Congo	*fē*		*fē*
Holoholo - NC:Bantu D	*-hyǎ*		*-hyǎ*
Kuria - NC:Bantu E	*oko-hya*	*oko-hya*	
Ngoreme - NC:Bantu E	*oko-hya*	*oko-hya*	
Bungu - NC:Bantu F	*kuku-pya*	*kuku-pya*	
Langi - NC:Bantu F	*ku-fya*		*fya*
Sukuma - NC:Bantu F	*(gU)pya*	*gUpya*	
Nyaturu - NC:Bantu F	*u-fya*	*u-fya*	
Nyamwezi - NC:Bantu F	*-pya*	*-pya*	*-pya*
Sumbwa - NC:Bantu F		*-hya*	*-hyahya*
Wungu - NC:Bantu F	*kuku-pya*	*kuku-pya*	
Mambwe - NC:Bantu F	*pya*		*pya*
Hehe - NC:Bantu G	*(uku)pya*	*(uku)pya*	
Kisi - NC:Bantu G	*ku-pya*	*ku-pya*	
Shashi - NC:Bantu G	*(oku)hya*	*(oku)hya*	
Wanji - NC:Bantu G	*(ku)pia*	*kupia*	
Asu - NC:Bantu G	*ʃa*		*ʃa*
Taveta - NC:Bantu G	*ʃa*		*ʃa*
Haya - NC:Bantu J	*(ku)hya*	*kuhya*	*-hya*
Ha - NC:Bantu J	*ugu-sha*	*ugu-sha*	
Nyambo - NC:Bantu J	*(ku)sya*	*(ku)sya*	
Nyankore - NC:Bantu J	*ku-sya*	*ku-sya*	
Shubi - NC:Bantu J	*gu-sha*	*gu-sha*	
Zinza - NC:Bantu J	*ku-sha*	*ku-sha*	
Kete - NC:Bantu L	*- ƚsw*		*- ƚsw*
Namwanga - NC:Bantu M		*uku-pya*	*-pya*
Nyiha - NC:Bantu M	*ku-pwa*	*ku-pwa*	
Rungu - NC:Bantu M	*pya*		*-pya*
Wanda - NC:Bantu M	*(uku)pya*	*(uku)pya*	
Chingoni - NC:Bantu N	*pya*		*pya*
Tswana - NC:Bantu S	*s’a*		*-s’a*

Table 35: ‘new’ = ‘fire; burnt’, ‘new’ = ‘ripe’ and ‘fire; burnt’ = ‘ripe’

This is clearly a case of homonymy rather than polysemy. Even for a non-specialist, these regular homophones allow us to establish sound correspondences between all these languages. Here are a few of them (Table 36):

Language	‘fire; be burnt’	‘ripe’	‘new’	C1
Proto-Bantu	*pí, pía*	*pí*	*pía, pái*	****pɪ***
Nyamwezi - NC: Bantu F	*-pya*	*-pya*	*-pya*	***py***
Nyiha - NC: Bantu M	*ku-pwa*	*ku-pwa*		***pw***
Holoholo - NC: Bantu D	*-hyǎ*		*-hyǎ*	***hy***
Haya - NC: Bantu J	*(ku)hya*	*kuhya*	*-hya*	***hy***
Rombo - NC: Bantu E	*-ia*		*-iâ*	***i***
Langi - NC: Bantu F	*ku-fya*		*fya*	***fy***
Asu - NC: Bantu G	*ʃa*		*ʃa*	***ʃ***
Ha - NC: Bantu J	*ugu-ʃa*	*ugu-ʃa*		***ʃ***
Nyankore - NC: Bantu J	*ku-sya*	*ku-sya*		***sy***
Kete - NC: Bantu L	*- ƚsw*		*- ƚsw*	***tsw***

Tswana - NC: Bantu S	*s'a*		*-s'a*	***s'***

Table 36: regular sound correspondences highlighted by regular homophony

It is easy to make sure that these correspondences are regular. As an experiment.,we can ask the reader to suggest what the Kete reflexes of the proto-Bantu forms **pɪkà* 'slave' and **pígò ~ píjò* 'kidney' should be? From the above table, the corresponding Kete words should begin with *tsw-*. And they do! Kete *tswîk* 'slave', *tswîc* 'kidney'.

The triple homophony 'new' = 'fire; be burnt' = 'ripe' has to be reconstructed for proto-Bantu (and it is indeed) but not for Benue-Congo, where it is absent. This can help understand the evolution of proto-Bantu from proto-Benue-Congo.

The next example of regular homophony is even more interesting, because it involves languages of various branches of Niger-Congo.

Our program retrieves the same Proto-Bantu form **tʊ́é* for the semantically unrelated meanings 'ashes' and 'ear'[12]. This homophony also manifests itself, for instance, in Okande (Bantu B) as *mó-tɔnì / mí-tɔnì* 'ashes' ~ *ɱ́-tónì / má-tónì* 'ear' or in Kete (Bantu L) as *-cw'* 'ashes' = 'ear'. Many actual Bantu languages have lost this homophony because in these languages the prefix of noun class 15 **ku-* in the word for 'ear' have merged with the lexical stem (Table 37):

Language	'ashes'	'ear'
Fipa - NC:Bantu F	*ii-t^{w}i*	*ii-ku-t^{w}i*
Wungu - NC:Bantu F	*i-t^{w}it^{w}i*	*i-ku-t^{w}i*
Mambwe - NC:Bantu F	*i-t^{w}i*	*i-ku-t^{w}i*
Rungu - NC:Bantu F	*t^{w}i*	*ku-t^{w}i*

Table 37: reflexes of the proto-Bantu homophony 'ashes' = 'ear'

This homophony is not restricted to Bantu. It can be found in Bantoid, Cross-River, Jukunoid and Nupoid, four other branches of Benue-Congo (Table 38).

Language	'ashes'	'ear'
Abanyom - NC:Benue-Congo:Bantoid	*à-tə̂ŋ*	*ɔ̀-tôŋ*
Balep - NC:Benue-Congo:Bantoid	*ǹ-dɔ̂n ~ bà-[r-]*	*ò-rôn*
Oro - NC:Benue-Congo:Cross River	*ń-tɔ́ŋ*	*ú-tɔ́ŋ*
Jibe - NC:Benue-Congo:Jukunoid	*sǫ́ǫ́*	*sǫ́ǫ̀*
Proto-Jukunoid	**tóŋ*	**tóŋ*
Gwari - NC:Benue-Congo:Nupoid	*ō-tnú*	*tnū-b^{w}à*

Table 38: other instances of 'ashes' = 'ear' in Benue-Congo

In other words, this homophony was already present in proto-Benue-Congo. It is attested in other families of Niger-Congo as well, although the forms may not be strictly identical (Table 39):

Language	'ashes'	'ear'
Kare - NC:Adamawa	*súù*	*súù*
Ɓali - NC:Adamawa	*tɛ̰́*	*tɛ́ / tɛ́rɔ̀*
Agni - NC:Kwa	*ǹzʊ̰́-ə̰̀*	*à-nzʊ̀-ɛ́*
Nzema - NC:Kwa	*ǹzɔ̰̀lɛ̰́*	*ǹzʊ̰̀ə́lɛ̀*
Nawuri - NC:Kwa	*ń-sɔ́*	*gʊ̀-sʊ́*

Table 39: 'ashes' = 'ear' in non-Benue-Congo Niger-Congo

[12] In both cases, the reconstructions are labeled as variants. The main reconstructions are *tó* 'ashes' and *tʊ́ì* 'ear', which does not affect the claims presented here. The data come from BLR3 online (http://www.africamuseum.be/research/discover/human_sciences/culture_society/blr/, consulted June 2018).

We can exclude the possibility that these coincidences are random. The similarity of forms cannot be explained semantically in this case. And this can only mean one thing: we discovered two Proto-Niger-Congo stems, close phonetically, and we found regular phonetic correspondences between various families of Niger-Congo, bypassing the stage of intermediate reconstructions in separate Niger-Congo families, groups and branches. Incidentally, this shared homophony proves the genetic relationship between all these languages, and therefore between the families to which they belong.

The said sound correspondences may now be confirmed by other non-homophonous forms, as for example the word for 'three' (Table 40):

Language	'ashes'	'ear'	'three'
Proto-Bantu	**tʊ́é*	**tʊ́é*	**tátʊ̀*
Balep - NC:Benue-Congo	*ǹ-dɔ̂n ~ bà-[r]*	*ò-rôn*	*bɛ́-rá*
Oro - NC:Benue-Congo	*ń-tɔ́ŋ*	*ú-tɔ́ŋ*	*ìtɛ́*
Jibe - NC:Benue-Congo	*sǫ́ǫ́*	*sǫ́ǫ̀*	*sàːr*
Kare - NC:Adamawa	*súù*	*súù*	*sáí*
Ɓali - NC:Adamawa	*tɛ́*	*tɛ́ ~ tɛ́rɔ̀*	*taat*
Agni - NC:Kwa	*ǹzʊ́g̰à*	*ànzʊ̀ɛ́*	*ńzg̰à, ńsĝ̰*
Nzema - NC:Kwa	*ǹzɔ̰̀lɛ́*	*ǹzʊ̰̀gɛ́lɛ̀*	*ǹsg̰̀*
Nawuri - NC:Kwa	*ń-sɔ́*	*gʊ̀-sʊ́*	*à-sá*

Table 40: A regular sound correspondence in Niger-Congo

Once the sound correspondence is found and cognate sets are defined, it becomes possible to look for more correspondences in related languages that don't show the homophony 'ashes' = 'ear'. These can be found not only in Benue-Congo, Adamawa or Kwa, but also in Bangime, an isolate Niger-Congo language (Table 41):

Language	'ashes'	'ear'	'three'
Eloyi - NC:Benue-Congo	*kú-lɔ̀*	*kɔ̄-lù*	*kō-lá*
Sambe - NC:Benue-Congo	*à-tɔ̄*	*(kū)-tûn*	*kà-tār*
Tupuri - NC:Adamawa	*sḡ̰ḡ̰gē*	*súgí*	*sùwàʔà*
Samba Leko - NC:Adamawa	*lā-tēm*	*túŋ*	*tōōrə̄*
Yoti - NC:Adamawa	*tìntá*	*tóo*	*táat*
Chumburung - NC:Kwa	*ń-sɔ́*	*kì-síbɔ́*	*à-sá*
Logba - NC:Kwa	*ntɔ́*	*òtswé*[13]	*ìtà*
Bangime - NC:isolate	*tíɥɛ̀*	*tàŋà*	*tààrù*

Table 41: more Niger-Congo cognates

Now, from these well-established sound correspondences, one may infer that proto-Niger-Congo showed the same initial consonant in the words for 'ashes', 'ear' and 'three'. And we may now add uncomplete series from other NC families (Table 42):

Language	'ashes'	'ear'	'three'
Moyobe - NC:Gur		*kù-tù*	*tàànì*
Nyun-Gunyaamoolo - NC:Atlantic		*bʊ-laaŋ*	*ha-lall*
Pre - NC:isolate	*tùɣùnì*		*tàːŋgɔ́*
Limba - NC:isolate	*fɔ-tɔ*		*bi-tat*
Dida (Lakota) - NC:Kru	*títɔ̄*		*tā*
Aizi - NC:Kru	*tʊtʊ*		*ita*
Ikposso - NC:Kwa	*ɔ́tīlì*	*ʊ́tʊ́ ~ átʊ́*	

[13] Here we postulate that **tue* > *tswe*, since affrication is common before closed vowels.

Table 42: partial correspondence sets in NC

To sum up, homophones found in various branches of NC was used to discover a robust set of sound correspondences throughout all the phylum, without the help of intermediate reconstructions. From a limited number of series, we arrived at a very high degree of confidence because the presence of the same homophonous pairs in many languages cannot be due to chance. In addition, we have found an evidence that some languages classified as being isolates, such as Bangime, belong to Niger-Congo.

Proto-Bantu has been reconstructed with two tones, H and L. But if an actual Bantu language has lost tones (as is the case for Swahili), or if the sources in the database just do not mention them (which is often the case), then it will lead to a number of segmentally homophonous pairs. For example, proto-Bantu **dà* 'belly' is segmentally homophonous with **dá* 'louse'. It is therefore no surprise if our database contains 29 sources that illustrate this homophony, all belonging to the Bantu stock (Table 43):

Language	Form	Meanings
Ikizu - NC:Bantu E	*enda*	'**belly**', 'pregnancy', '**louse**'
Ikoma - NC:Bantu E	*anda*	'**belly**', 'pregnancy', '**louse**'
Bende - NC:Bantu F	*nda*	'**belly**', 'womb', 'pregnancy', '**louse**'
Fipa - NC:Bantu F	*iinda*	'**belly**', 'womb', 'pregnancy', '**louse**'
Hangaza - NC:Bantu J	*inda*	'**belly**', 'womb', 'pregnancy', '**louse**'
Zinza - NC:Bantu J	*eenda*	'**belly**', '**louse**'
etc.		

Table 43: 'belly' = 'louse' in Bantu

The same thing happens with the reflexes of proto-Bantu *cí 'ground; country; underneath' and *cì 'to grind' but this case is more interesting, because this homophony is also found outside Bantu. In the Banda languages, a branch of the Ubangi family of Niger-Congo, *ʃī* means both 'to bury' and 'to grind'. This form is very close to the Bantu forms. Moreover, a similar stem is found in Mande (whose Niger-Congo membership is disputed): Bambara *sí* 'to crush, to grind', Mandinka *sîi* 'to grind', Niokolo Maninka *sii* 'to crush, to grind'. Thus, thanks to the fact that some authors neglected to note tones, our automated search came up with two probable Niger-Congo lexical roots that were phonetically close to each other: 'to bury' and 'to grind'.

Another example of widespread Niger-Congo homophony is given by the pair 'to bite' = 'to extinguish'. This homophony is strict in a few languages from 3 NC families: Adamawa, Kwa and Atlantic (Table 44):

Language	'to bite'	'to extinguish'
Tula - NC:Adamawa	*dùm*	*dùm*
Chumburung - NC:Kwa	*dùŋ*	*dùŋ*
Krache - NC:Kwa	*dùŋ*	*dùŋ*
Nawuri - NC:Kwa	*dùŋ*	*dùŋ*
Gonja - NC:Kwa	*dùŋ*	*dùŋ*
Bijogo - NC:Atlantic	*num*	*num*

Table 44: 'to bite' = 'to extinguish'

This set of homophones looks limited but it still suggests that in Proto-Niger-Congo these meanings were expressed with either identical or very similar forms. Actually, whenever these two roots have been reconstructed, the forms are very similar (Table 45):

proto language	'to bite'	'to extinguish'
Proto-Bantu - NC :Benue-Congo	*dʊ́m*	*dím*

Proto-Lower-Cross - NC :Benue-Congo	*lóm*	*nímɛ́*
Proto-Upper-Cross - NC :Benue-Congo	*dómà*	*dímà*
Proto-Jukunoid - NC :Benue-Congo	*dʷàm*	*dim*

Table 45: reconstructions of 'to bite' and 'to extinguish' in Niger-Congo branches

All the above reconstructions pertain to Benue-Congo, and show a contrast between the vowel of 'to bite' and that of 'to extinguish'. The same contrast can be found throughout nearly all branches of Niger-Congo, although reflexes of both roots are found only in a few languages (Table 46):

Language	'to bite'	'to extinguish'
Obolo-Andoni - NC:Benue-Congo	*lóm*	*nííŋ*
Kom - NC:Benue-Congo	*lōm*	*lʊ̄msə́*
Abanyom - NC:Benue-Congo	*lôm*	*lɨ́mî*
Joola fogny - NC:Atlantic	*ɛrʊm*	
Basari - NC:Atlantic		*a-lɔ́mî / ɗɔ́mî*
Palor - NC:Atlantic	*ɗob*	*ʃim*
Landuma - NC:Mel		*dim*
Sherbro - NC:Mel		*rim*
Limba - NC:isolate	*tɔma* 'to eat'	*domina*
Samba leko - NC:Adamawa	*lùm*	
Dii - NC:Adamawa		*nim*
Tupuri - NC:Adamawa	*lɔ̀ŋ-gɛ̀*	
Day - NC:Adamawa	*ním*	
Kare - NC:Adamawa	*núŋ*	
Yingilum - NC:Adamawa	*numtɛ̄*	
Mbandja - NC:Ubangi	*nò*	
Mundu - NC:Ubangi	*nìmà ~ ɲìmà*	
Linda - NC:Ubangi	*lō*	*rɬ*
Kabiyè - NC:Gur	*ɖʊ́m̀*	*ɖím̀*
Pana - NC:Gur	*dèm*	*dímésé*
Bambara - NC:Mande	*dṹ* 'to eat'	
Mandinka - NC:Mande	*dómò* 'to eat'	
Bassa - NC:Kru	*nūmū*	
Mombo - NC:Dogon	*núngɛ́*	
Ị̱bani̱ - NC:Ijo	*ónwí*	

Table 46: 'to bite' and 'to extinguish' in Niger-Congo

The preliminary reconstructions for 'to bite' and 'to extinguish' may be posited as **dʊm* and **dɪm* respectively. In some languages the roots converged (Tula, Gonja, Bijogo…) while in some others they diverged (Linda, Obolo-Andoni, Palor). But in most cases one of the roots was replaced. However, the presence of homophones in the database allowed us to establish the series exemplified in Table 46, which in turn enabled us to set regular sound correpondances. Given that chance resemblences are extremely unlikely, it is safe to posit that, for example, *n-* in Bijogo (Atlantic) corresponds to *d-* in Gonja (Kwa) and to *l-* in Tupuri (Adamawa). It should be emphasized once more that this result has been obtained without any intermediate reconstructions of, say, proto-Adamawa or proto-Kwa.

5 Discussion

5.1 How “homophonic” is a language?

In the RefLex database, we find around 100,000 homophones, i.e., word forms with several meanings. Naturally, some sources have more homophones than others. What are the parameters that influence the number of homophones in a given language? tone? dominant syllable structure? phonemic inventories? average word length? After having played with all these hypotheses, we noticed that different sources for a given language may vary a lot when it comes to the number of homophones. In Bambara for instance (NC:Mande), the old 300-wordlist of S. Koelle (1854) contains only 2% of homophones, whereas the huge dictionary (23,170 entries) by G. Dumestre (2011) has about 20%. Therefore, it seems that the number of homophones is not a property of a language, but rather a property of a lexicographical source, depending on how the author chooses to encode the meaning of each form. But we also noticed that the more entries a source contains, the more homophones it has. In Table 47 below, sources are ranked by their number of entries and for each category the average number of homophones is given. As the size of the sources decreases, the number of homophones decreases as well.

# of words	>7000	5000-7000	1000-5000	700-1000	500-700	300-500	19-300	Total[14]
# of sources	20	20	255	103	101	156	543	1198
homophones (%)	13%	12%	9%	7%	6%	4%	3%	

Table 47: correlation between the size of lexical sources and their percentage of homophony

These are the average numbers of homophones. There can be important variations if one looks at particular sources. For example, the Sereer dictionary (NC:Atlantic) of Crétois (1973-7) has 21,623 entries but only 7% homophones. On the other hand, in a very short wordlist of Abure (NC:Kwa, Burmeister 1983), there are only 57 words but 12 homophones (21%). But despite these local variations, the general trend is quite robust.

Is this statistical trend a feature of the individual sources or a feature of African lexicography in general? Behind the numbers, one can see the various strategies used by the authors to describe meanings. If an author is very precise in describing the meaning of a given form, there will be more homophones, but also more entries in the dictionary. On the other hand, authors who only gather short, Swadesh-type wordlists generally don’t bother to explore the full ranges of meaning of the words they collect. This results in less homophony. These phenomena are characteristic of lexicographical practices, which led some colleagues to suggest the term “homograph” instead of “homophone”. But despite all these facts, we believe that the lexicographical information contained in as big a database as RefLex is, as a whole, more “objective” than the individual, subjective approaches of the authors of each dictionary. Therefore, we consider our “big data” approach teaches us not only about sources, but also about languages,.

5.2 A typology of regular homophony

As we have shown, homophones may be very different in nature. We are now in a position to list every possible type of homophony:

1) General, regular polysemy, which does not depend on language relatedness.

2) Local, regular polysemy, which characterizes a group of genetically related languages or a specific geographical area.

3) Local, regular homonymy which caracterizes a group of genetically related languages, and which is the consequence of homonymy (or near homonymy) in the corresponding proto-language.

[14] The database contains 1289 sources, as of June 2018. The total here is lower because 91 sources (mostly very short wordlists) show no homophones at all.

4) Noise. Noise includes: i) grammatical polysemy (*a work / to work*); ii) the overly-detailed description of a meaning ('to enter' / 'to go in' / 'to come in'); iii) chance regular homonymy (Samba Leko - NC:Adamawa *mə́* '1st pers. sg.' = 'to hide', Dan - NC:Mande *bǧ́* '1st pers. sg.' = 'to hide'); iv) "technical" homonymy, for instance cases where the author included the same word several times in the original source.

Ideally, each of these four cases could be easily recognized. If a homophony is attested in different (unrelated) language families, then case 1 is the most likely, i.e. regular polysemy ('arm' = 'hand', 'milk' = 'breast' and the like). If it is restricted to a single language family and occurs in different languages with different word roots, then it is case 2, i.e. local polysemy (see examples below). Finally, if roots are similar and no semantic link is found, then it is case 3, i.e. local homonymy, which is only possible among related languages. Of course, classifying regular homophones is not always that simple.

Let us go back to one of the examples cited above (§3.2.3), namely 'tooth' = 'name'. The fact that it is attested not only in Niger-Congo languages but also in South Bauchi (AA:Chadic) - a branch of Afroasiatic - should lead us to consider it an instance of regular polysemy (case 1). But there is no simple way of putting these two meanings together. Moreover, this homophony does not occur in every Chadic language - only in a small cluster of languages that are in contact with languages from various Niger-Congo branches. To explain this, there are two competing hypotheses, each one starting with the fact that there were two close proto-Niger-Congo roots that merged in several daughter languages. The first hypothesis is that Proto-South-Bauchi could have borrowed the homophones from one of these languages. The second hypothesis builds on the fact that proto-Afroasiatic roots for 'name' and 'tooth' are reconstructed as **(ʔi-)sim-* and **sin-* respectively (Militarev & Stolbova 2007). Interestingly, in Chadic, only the South-Bauchi languages seem to have reflexes of both proto-forms (see Table 16 above), perhaps under the influence of neighbouring NC languages. We don't have any reliable reconstructions for Niger-Congo yet, but from the actual forms we may hypothesize that 'name' and 'tooth' were roughly like ***ʔIN* and ***NIN* respetively. So, it just so happens that the forms were close in both protolanguages and that some of the daughter languages sporadically merged them.

To assess the importance of each one of the four distinct cases of regular homophony, we conducted an analysis of the hundred meanings in the Swadesh list (removing 'this', 'that' and 'no'). Cases of regular polysemy for each notion are listed in the appendix.

In average, there are about 10 polysemic links for every meaning of the list. But the actual number of links varies considerably from one meaning to another. The most polysemic meaning are presented in Table 48, with the most frequent polysemic links on the left, and the least frequent on the right:

Meaning	Nb of regular polysemic links	Meaning	Nb of regular polysemic links
'to say'	34	'smoke'	4
'big'	32	'fish'	4
'good'	27	'tail'	4
'mouth'	26	'to fly'	4
'skin'	25	'knee'	3
'small'	23	'two'	2
'to burn'	22	'egg'	2
'seed'	22	'louse'	2
'ground'	22	'star'	1
'to eat'	21	'dog'	0

Table 48: the 10 most polysemic and the 10 least polysemic notions (from the Swadesh 100-list)

Before carrying on with the analysis of these data, a few problems have to be addressed. First, the program that extracts homophonic relations from RefLex relies on the French translations, and that may yield slightly different results from English-based results. For instance, for the item 'mouth' (French *bouche*), some of the relations would be different in English: *embouchure* ('mouth' of a river) translates as 'mouth' and the French relation *bouche* <> *embouchure* would become *mouth* <> *mouth*; therefore it would not be counted as a relation. *Bout* and *fin* both translate as 'end' so their scores should be added, etc. Thus, we have decided to provide the meanings in French as we found them in the database in the appendix.

Second, some of the meanings could be merged (even in French), either for semantic reasons (as for instance *beaucoup* 'many, much' and *abondant* 'abundant') or for syntactic reasons (for the item 'big', the links 'old' and 'to be old' are counted separately), but this cannot be easily automatized. Therefore, the tables in the appendix give a general overview of the polysemic links that can be extracted from RefLex but the frequency given in brackets should not be taken as absolute.

Still, it appears that some notions included in the basic vocabulary have few polysemic relations. 'Dog' and 'star' are only related to 'poor' and 'firefly' respectively. This latter polysemy is attested in Day (NC:Adamawa), in proto-Bantu and in three Bantu A languages, but interestingly, the roots involved in the actual Bantu languages are different from the one reconstructed in proto-Bantu (Table 49). It means that a polysemic link may be preserved even if the word is renewed.

Language	'star' = 'firefly'
Day - NC:Adamawa	*mbɔ́l mbīŋā*
proto-Bantu - NC:Bantu	**ɲéɲèdí*
Koonzime - NC:Bantu A	*cwícwén*
Kwakum - NC:Bantu A	*ʃɔ̀ʃɔ́n*
Shiwa - NC:Bantu A	*yákū*

Table 49: 'star' = 'firefly'

The forms 'star' and 'dog' have no or few polysemic links but they may be involved in local homophony (our case 3): 'star' = 'to pound' in five Dogon languages, 'star' = 'feather' in some Bantu C languages, 'dog' = 'twenty' in the Nupoid group of Benue-Congo. Local polysemy (our case 2) is also attested with these notions: 'star' = 'hail' with different roots in the Mel branch of NC (Kisi *lùé*, Temne *k-ɔs*); 'dog' = 'poor' in Bantu (zones BCHK). And finally, chance homonymy (case 4) seems to exist too: Bakwe (NC:Kru) *fɛ̀* 'dog' = 'to do', proto-Jukunoid *bu* 'dog' = 'to do', but in the latter case the tone are not reconstructed and might have been different.

Now let us look in more detail at one of the highly polysemic items of the Swadesh list, namely 'mouth'. The automated search yields 63 different cases of regular *(*i.e. attested in 2 or more languages) homophony. Among those, 5 are cases of local homonymy: 'mouth' = 'cow' = 'to give birth' (Moru-Madi - NS:Central Sudanic); 'mouth' = 'dry season' (Yoruba lects - NC:Defoid), etc.; 18 links are to be counted as noise; 5 homophones can be counted as local polysemy (for instance, 'mouth' = 'tip' in Bantu J (NC)). We are left with 35 cases of regular polysemy (case 1). The meanings involved may be classified in broad semantic domains:

A: 'language'; 'speech'; 'word'; 'message'; 'song'; 'curse'; 'to keep silent'; 'to yawn'.
B: 'to drink'; 'mouthful'; 'pot'.
C: 'lip'; 'beak'; 'palate'; 'jaw bone'; 'maw'.
D: 'front'; 'on'; 'place'.
E: 'edge'; 'opening'; 'beginning'; 'end'; 'door'; 'entrance'; 'threshold'; 'limit'; 'well (n.)'.
F: 'blade'; 'sharp'; 'to shave'.
G: 'number'; 'to measure'; 'price'; 'to cost'.

Groups A and B are related to the functions of the mouth. In Africa the link between 'mouth' and 'speech' is very widespread. But if the mouth speaks, it also drinks (curiously there is no case of 'mouth' = 'to eat'). Group C includes body parts related to the mouth in one or the other way. Group D is based on the position of the mouth. Group E combines the position and the shape of the mouth. Group F may be regarded as a specification of group E. It is presented in details in Table 50:

Language	Form	Meanings
Afar - AA:Cushitic	*áf ~ áfa*	'mouth' , 'sharp', 'blade'
Limba - NC:isolate	*foti*	'mouth', 'blade', 'sharp'
Bobo Madare - NC:Mande	*dō*	'mouth', 'edge', 'sharp', 'number'
Lele - NC:Mande	*dá ~ dà*	'mouth', 'blade', 'door'
Bwamu - NC:Gur	*ɲṵ̄*	'mouth', 'sharp', 'limit'
Jamsay - NC:Dogon	*káá*	'mouth', 'edge', 'to shave'
Yoruba - NC:Benue-Congo	*ēnū*	'mouth', 'sharp'
Bijogo - NC:Atlantic	*kana*	'beak', 'mouth', 'end', 'door', 'blade'

Table 50: a group of related semantic links from 'mouth'

Finally, group G looks more arbitrary than the others, but the distribution shows that these are not cases of homonymy. In fact, this case of general polysemy is attested in four different branches of the Niger-Congo phylum (Table 51):

Language	Form	Meanings
Bobo Madare - NC:Mande	dō ~ dā	'mouth', 'edge', 'sharp', '**number**'
Dan - NC:Mande	dǐ	'mouth', 'opening', '**number**'
Bokobaru - NC:Mande	lɛ́	'mouth', 'opening', 'fine', 'edge', '**to measure**' ...
Samo-Maya - NC:Mande	lɩ	'mouth', 'song', '**number**', 'edge', 'limit'
Jalonke - NC:Mande	dɛ́ɛ́	'door', 'mouth', 'opening', 'well (n.)', '**price**'
Mandinka - NC:Mande	dáá	'mouth', 'opening', 'edge', 'place', 'beginning', '**price**' ...
Maninka West - NC:Mande	dàà	'mouth', '**price**', 'to weave'
Chumburung - NC:Kwa	ká-nɔ́	'mouth', '**number**'
Krache - NC:Kwa	kɔ́-nɔ́	'mouth', '**number**'
Nawuri - NC:Kwa	gɔ́-nɔ́	'mouth', '**number**'
Gonja - NC:Kwa	kɔ́-nɔ́	'mouth', '**number**'
Tene Kan - NC:Dogon	ká̱	'mouth', '**number**', '**price**', '**to measure**', '**to cost**'
Toro Tegu - NC:Dogon	ká	'mouth', 'edge', '**to measure**'
Jamsay - NC:Dogon	káá	'mouth', 'edge', 'to shave', '**to measure**'
Ben Tey - NC:Dogon	mòó	'mouth', 'edge', '**to measure**'
Curama - NC:Gur	nu̱ɔ̱	'mouth', 'edge', 'opening', 'lineage', '**number of**'

Table 51: the polysemy of 'mouth': group G

As a whole, our approach, based on regular homophony, has shown a great effectiveness. Once the "noise" is removed, each and every regular homophone gives precious information, be it in semantics or in historical phonetics. In diachronic semantics, every case of polysemy broadens the basis for semantic reconstruction as well as our arguments for finding new cognates. In synchronic semantic typology, the availability of a complete list of quantified polysemic links for a given notion allows us to work with multidimensional semantic fields instead of individual links, as was shown above with the example of 'mouth'.

5.3 A powerful tool for reconstruction

In diachronic phonology, every case of regular homophony gives us a very strong basis for the establishment of regular sound correspondences. Such correspondences may even be set up

between genetically distant languages, without having to rely on intermediate reconstructions. The correspondences discovered with this method need not be treated as hypothetical, since they are statistically proven. But let us examine in more detail this heuristic aspect of our approach. Table 52 shows regular sound correspondences between two cognate sets obtained by comparing lexicons from two NC branches of West Africa, Atlantic and Mel:

Language	'onion'	'to answer'
Jaad - NC:Atlantic	*jabɛ*	*jaab-iŋ-*
Sereer - NC:Atlantic	*jaba*	*jab-oh*
Nalu - NC:Atlantic	*yabɛ*	*yaab-ŋah*
Landuma - NC:Mel	*jʌbʌ*	*jabi*
Temne - NC:Mel	*kə-yaba*	*a-yabi*

Table 52: 'onion' and 'to answer' in Atlantic and Mel

How valid are these correspondences? Actually, it is not difficult to show that these forms were all borrowed from West Mande languages, with which Atlantic and Mel languages have been in contact for centuries. Here are a few Mande forms (Table 53):

Language	'onion'	'to answer'
Jalonke	*jabɛ*	*jáabi*
Susu	*yɛbɛ, yabɛ*	*yaabi*
Mandinka	*jaba*	*jaabi*
Maninka (Niokolo)	*jába*	*jáabi*
Bambara	*jàba*	*jáabi*
Soninke	*jàbâ*	*jáabì*

Table 53: 'onion' and 'to answer' in West Mande

Now let us use the approach presented in this paper. Only among the Upper-Cross languages (NC:Benue-Congo), do we find the regular homophony 'hair' = 'root' = 'vein' (Table 54):

Language	'hair'	'root'	'vein'
Doko-Uyanga	*diŋ*	*diŋ*	*diŋ*
Iyoniyon	*ri-tin*	*ri-tin*	
Kohumono	*sìn*	*sìn*	
Kukele	*ttílí*	*ttílí*	
Lubila	*liŋ*	*liŋ*	*liŋ*

Table 54: 'hair' = 'root' = 'vein' in Upper-Cross languages

The polysemy 'root' = 'vein' is widely attested in Africa (44 languages in our database), and it is undoubtedly a case of general regular polysemy. However, the homophony 'hair' = 'root' is unique and limited to Upper-Cross languages. Since the forms are phonetically comparable, we can consider the correspondence Doko *d-* / Iyoniyon *t-* / Kohumono *s-* / Kukele *tt-* / Lubila *l-* as valid. Unlike pseudo-cognates as 'onion' and 'to answer' above, homonymic cognates give us a true phonetic correspondence since the possibility of borrowing is minimal here. If we were to consider this as a case of a borrowing, then we would have to admit that languages can borrow not only words and their meanings, but also complete sets of homonyms. Even if such situations may occur, they must be marginal and may be discarded. Thus, our method is also a way to prove that a cognate set does not include borrowings.

The final result of Niger-Congo diachronic phonetics could be a huge matrix of about 1,500 rows (for languages and proto-languages) and a few dozen columns for NC proto-phonemes. The cells of the table would be filled with the corresponding reflexes in all the languages and proto-languages. Naturally, there is little hope that this matrix will be entirely filled up in the next decades. But we are

confident that our approach offers an unexpected possibility to contribute to the filling of various fragments of the matrix without having to wait for the progress of Niger-Congo comparative-historical linguistics. In other words, working with regular homophones is a win-win game : we are guaranteed to discover either semantic parallels, or regular sound correspondences of the best quality.

6 Conclusion

Building on a systematic survey of homophones (defined as including both polysemic and homonymic forms) in a very large lexical database of African languages[15], we have developped an approach which is powerful and promising for both lexical typology and historical linguistics. In lexical typology, the actual frequencies of polysemic links allow us to assess the semantic extension of a given concept on a more objective basis than intuition alone. In historical linguistics, our approach proves fruitful in two directions: First, regularly attested polysemic links give licence to consider as cognates forms whose meaning may differ (see §3.4.1). Second, regular homonymy (i.e. similar forms with unrelated meanings), since it can hardly be attributed to chance, constitutes a very strong piece of evidence not only for the relatedness of the concerned languages but also for the regularity of the sound correspondences involved (see §4).

7 References

Asher, R. E. (Editor-in-Chief) & J. M. Y. Simpson (Coordinating Editor). 1994. *The Encyclopedia of Language and Linguistics*. Oxford, New York: Pergamon Press.

Dumestre, Gérard. 2011. *Dictionnaire bambara-français - suivi d'un index abrégé français-bambara*. Paris : Karthala.

Gévaudan, Paul, Peter Koch & Antonia Neu. 2003. Hundert Jahre nach Zauner: Die romanischen Namen der Körperteile im DECOLAR. *Romanische Forschungen* 115(1), pp. 1–27.

Koelle, Sigismund Wilhelm. 1963 [1854]. *Polyglotta africana, or a comparative vocabulary of nearly three hundred words and phrases in more than one hundred distinct African languages*. London : Church Missionary House.

[Kuznetsova] Кузнецова, Ариадна Ивановна; Татьяна Федоровна Ефремова. 1986. *Словарь морфем русского языка*. М., “Русский язык”.

List, Johann-Mattis, Anselm Terhalle and Matthias Urban. 2013. *Using Network Approaches to Enhance the Analysis of Cross-Linguistic Polysemies*. Proceedings of the 10th International Conference on Computational Semantics -- Short Papers. Association for Computational Linguistics pp. 347-353.

Matthews, Peter H. 2007. *The Concise Oxford Dictionary of Linguistics*. Oxford, New York: Oxford University Press. Second Edition.

Militarev, Alexander & Olga Stolbova. 2007. *Afroasiatic etymology*. http://starling.rinet.ru, accessed June 2018.

Moñino, Yves. 1995. *Le proto-gbaya: essai de linguistique comparative historique sur vingt-et-une langues d’Afrique centrale*. Paris : Peeters, SELAF.

[15] By the time we finished writing this paper, we became aware of a comparable experiment (Pericliev 2015). In this article, the author lists all the ‘colexifications’ (what we call here ‘homophones’) among the ASJP database (Wichmann et al. 2013). However, Pericliev’s approach differs from ours in several respects. First, he considers basic vocabulary only, since it is what the ASJP database contains. Second, the regularity of colexifications is mostly used to establish or confirm genetic relationships between languages or language groups.

Pericliev, Vladimir. 2015. On Colexification among Basic Vocabulary. *Journal of Universal Language* 16(2), pp. 63-93.

Pozdniakov, Konstantin & Guillaume Segerer. In press. 'A Genealogical classification of Atlantic languages' in Luepke, Friederike (ed). *The Oxford guide to the Atlantic languages of West Africa*. Oxford : Oxford University Press.

Segerer G., Flavier S., 2011-2018. *RefLex: Reference Lexicon of Africa*, Version 1.1. Paris, Lyon. http://reflex.cnrs.fr/.

Wichmann, Søren, Eric W. Holman, and Cecil H. Brown (eds.). 2013. The ASJP Database (version 16).

Zalizniak, Anna. 2008. A catalogue of semantic shifts: towards a typology of semantic derivation. In Martine Vanhove (ed.), *From Polysemy to Semantic Change: Towards a typology of lexical semantic associations*. Amsterdam & Philadelphia: John Benjamins, pp. 217-238.

Zalizniak, Anna A., Maria Bulakh, Dmitrij Ganenkov, Ilya Gruntov, Timur Maisak & Maxim Russo. 2012. The catalogue of semantic shifts as a database for lexical semantic typology. *Linguistics* 50(3), pp. 633-669.

8 Appendix

The regular polysemy in the 100-words Swadesh list in African languages

In the tables below, we list all the words from the Swadesh 100-list and their polysemic links. Three items ('this', 'that' and 'not') have been removed. The items have been sorted by decreasing number of polysemic relations, then alphabetically. Items are given in French and English but the related meanings are given only in French, since French was used to retrieve them from the database (for a more elaborated justification, see § 5.2). For each table, the first line gives the number of polysemic relations; the second line contains the list items; the third line shows the number of sources which have at least one homophone for the corresponding item. Then come the related meanings (with the number of relevant sources in brackets). We show only meanings that are attested at least in 3 sources; as for the relations attested in 2 sources only, we include the most interesting cases. We have subjectively removed all links that we considered homonymic (i.e. not polysemic). As much as possible, similar meanings rendered by different French words have been merged (ex. *médicament* 'medicine' and *remède '*medicine, cure*'* merged as *médicament* under the item *feuille* 'leaf').

34	30	26	25	23
dire 'to say'	grand 'big'	bouche 'mouth'	peau 'skin'	bon 'good'
279 sources	*456 sources*	*338 sources*	*306 sources*	*288 sources*
parler (183)	gros (155)	lèvre (84)	écorce (122)	beau (194)
raconter (21)	puissant (121)	bord (23)	cuir (42)	bien (83)
faire (12)	long (98)	bec (19)	corps (41)	joli (34)
expliquer (12)	large (27)	ouverture (19)	livre (16)	bon goût (17)
appeler (9)	gras (25)	langue (17)	papier (13)	agréable (16)
compter (9)	haut (25)	porte (12)	chaussure (11)	doux (14)
informer (8)	épais (22)	bout (12)	écaille (8)	gentil (14)
annoncer (8)	vieux (18)	nombre (9)	fourrure (7)	propre (12)
sortir (8)	loin (17)	entrée (9)	ceinture (6)	sentir bon (11)
parole (7)	chef (15)	langage (8)	vêtement (6)	facile (11)
tomber (7)	mère (10)	parole (7)	lettre (6)	sucré (8)
mordre (7)	fort (10)	mesurer (6)	croûte (5)	beaucoup (5)
prononcer (6)	père (10)	limite (5)	chair (5)	plaire (5)
avertir (6)	être gros (9)	gueule (5)	gousse (5)	mou (5)
nommer (5)	important (9)	seuil (5)	pelure (4)	beauté (5)
être (4)	beaucoup (9)	pointe (4)	chapeau (4)	abondant (5)
suivre (4)	mûr (8)	couvrir (4)	coquille (3)	délicieux (4)
chanter (4)	énorme (8)	boire (4)	pagne (3)	correct (4)
gronder (4)	aîné (6)	tranchant (4)	carapace (3)	venir (4)
poser (4)	épais (5)	prix (4)	natte (3)	généreux (4)
répondre (4)	maître (5)	raser (3)	viande (3)	salé (4)
souffler (4)	force (5)	embouchure (3)	carapace (3)	heureux (3)
promettre (4)	frère aîné (4)	lame (3)	sac (3)	d'accord (3)
que (4)	vaste (4)	devant (3)	cocon (2)	
langage (3)	nombreux (4)	mot (3)	coquillage (2)	
couper (3)	adulte (4)	fin (3)		
se disputer (3)	âgé (3)			
comme (3)	croître (3)			
montrer (3)	roi (3)			
envoyer (3)	dur (3)			
finir (3)				
enseigner (3)				
avoir (3)				
résonner (3)				

23	22	22	22	21
petit 'small'	brûler 'to burn'	graine 'seed'	terre 'earth, soil'	manger 'to eat'
284 sources	*416 sources*	*187 sources*	*290 sources*	*291 sources*
peu (94)	rôtir (63)	semence (50)	sol (155)	prendre (100)
jeune (38)	flamber (38)	fruit (32)	pays (70)	mordre (24)
court (36)	mûrir (18)	noyau (15)	argile (24)	nourriture (19)
enfant (30)	feu (15)	enfant (15)	région (23)	croquer (17)
mince (21)	cuire (13)	oeil (14)	sable (20)	mâcher (17)
étroit (13)	chaud (12)	os (6)	monde (17)	ronger (14)
fruit (9)	allumer (9)	rein (5)	boue (11)	sucer (9)
bas (6)	chauffer (8)	grain (5)	poussière (10)	piquer (6)
maigre (6)	frire (7)	espèce (5)	parcelle (9)	gagner (6)
fils (6)	piquer (6)	testicule (5)	sous (9)	repas (6)
peu nombreux (6)	briller (6)	céréale (4)	village (9)	dépenser (5)
enfance (5)	manger (6)	pépin (4)	en bas (6)	tromper (5)
bébé (5)	mûr (4)	comprimé (3)	bas (6)	boire (5)
minuscule (5)	incendie (4)	germe (3)	banco (5)	consommer (5)
être court (4)	gratter (4)	noix (3)	champ (5)	dévorer (5)
menu (4)	bouillir (3)	clitoris (3)	endroit (5)	avaler (4)
frère cadet (4)	être chaud (3)	perle (2)	terrain (4)	se noyer (3)
humilier (4)	démanger (3)	clan (2)	mortier (3)	brouter (3)
bientôt (3)	cultiver (3)	race (2)	dessous (3)	brûler (3)
garçon (3)	blesser (3)	parent (2)	continent (3)	lécher (3)
nain (3)	allumé (3)	fils (2)	saleté (3)	droite (3)
léger (3)	faire sécher (3)	tribu (2)	ville (3)	
cadet (2)				

20	18	18	18	18
arbre 'tree'	froid 'cold'	pied 'foot'	tête 'head'	tuer 'to kill'
343 sources	*261 sources*	*381 sources*	*213 sources*	*350 sources*
bois (74)	mouillé (72)	jambe (336)	sur (24)	mourir (108)
bâton (48)	frais (27)	plante du pied (24)	sommet (16)	frapper (47)
médicament (29)	vent (19)	patte (18)	igname (9)	éteindre (29)
bois à brûler (18)	humide (13)	talon (12)	dessus (6)	battre (11)
médecine (17)	fièvre (12)	fois (11)	crâne (6)	blesser (8)
plante (14)	lent (11)	trace (9)	chef (6)	couper (7)
tronc (9)	calme (10)	roue (9)	sens (5)	abattre (animal) (5)
fruit (8)	paludisme (6)	sabot (7)	caractère (4)	abattre (5)
remède (8)	doux (6)	cuisse (6)	cause (4)	piler (5)
tronc d'arbre (8)	mou (5)	empreinte (5)	poignée (4)	jouer (4)
poison (7)	fraîcheur (5)	racine (4)	au-dessus (4)	détruire (3)
manche (6)	tranquille (5)	branche (3)	cheveu (3)	tomber (3)
tige (6)	harmattan (4)	traces (3)	haut (3)	casser (3)
canne (6)	humidité (4)	trace de pas (3)	personne (3)	récolter (3)
fétiche (5)	paix (3)	main (3)	bout (3)	effacer (3)
pilon (5)	saison froide (3)	fesses (2)	groupe (2)	lancer (3)
piquet (4)	faible (3)	base (2)	homme (2)	chasser (3)
perche (4)	trembler (3)	pas (2)	pointe (flèche) (2)	assassiner (3)
poutre (3)				

branche (3)				
17	16	16	15	15
donner 'to give'	écorce 'bark'	racine 'root'	beaucoup 'many'	homme 'man'
279 sources	*190 sources*	*165 sources*	*178 sources*	*353 sources*
offrir (57)	peau (92)	veine (44)	nombreux (42)	mâle (82)
prendre (24)	peau (fruit)	dent (14)	très (29)	mari (70)
mettre (13)	écaille (15)	nom (14)	plein (26)	esclave (48)
mordre (12)	coquille (12)	souche (5)	bien (14)	fils (23)
jeter (9)	pelure (9)	jambe (5)	remplir (14)	personne (22)
envoyer (9)	croûte (7)	tendon (5)	trop (13)	garçon (20)
rendre (9)	enveloppe (6)	nerf (5)	grand (9)	époux (8)
couler (6)	ongle (5)	pied (4)	plusieurs (9)	corps (7)
servir (5)	carapace (4)	médecine (3)	gros (9)	mariage (5)
vendre (5)	gousse (4)	arbre (3)	abondant (7)	peau (5)
poser (5)	vêtement (3)	médicament (3)	longtemps (5)	quelqu'un (4)
aider (5)	coque d'oeuf (3)	base (3)	tout (4)	divorce (4)
porter (5)	croûte (plaie)	tronc (3)	souvent (4)	chef (4)
passer (4)	griffe (2)	tubercule (3)	tas (4)	bouche (4)
nourrir (4)	cuir (2)	origine (3)	complètement (4)	beau-frère (3)
permettre (4)	coquillage (2)	source (2)		
cadeau (3)				

15	15	15	14	14
jaune 'yellow'	soleil 'sun'	ventre 'belly'	chaud 'hot'	main 'hand'
70 sources	*302 sources*	*294 sources*	*162 sources*	*477 sources*
rouge (14)	jour (113)	grossesse (85)	feu (35)	bras (390)
brun (8)	journée (13)	utérus (38)	brûler (16)	paume (54)
farine de néré (7)	Dieu (12)	dans (32)	chauffer (14)	avant-bras (14)
vert (7)	temps (11)	intestins (24)	rapide (12)	paume (11)
orange (5)	ciel (10)	estomac (24)	bouillir (10)	branche (10)
maïs (5)	lumière (9)	intérieur (14)	tiède (5)	doigt (9)
marron (5)	dieu (9)	intestin (6)	rouge (5)	poignée (9)
arbre sp. (4)	époque (6)	enceinte (5)	fiévreux (5)	cinq (8)
mûr (4)	est (6)	foetus (4)	transpirer (4)	patte (7)
plante sp. (3)	chaleur (4)	abdomen (4)	difficile (4)	poing (6)
bleu (2)	midi (4)	fond (3)	mûr (4)	manche (3)
fruit sp. (2)	après-midi (3)	dedans (3)	piquant (3)	griffe (3)
fauve (2)	fois (3)	entrailles (3)	vite (3)	pied (3)
pâle (2)	saison sèche (2)	boyaux (2)	sueur (3)	aile (2)
fruit de néré (2)	heure (2)	milieu (2)		

14	14	13	13	13
personne 'person'	vert 'green'	boire 'to drink'	feuille 'leaf'	pluie 'rain'
168 sources	*104 sources*	*178 sources*	*228 sources*	*277 sources*
gens (27)	non mûr (35)	fumer (58)	oreille (32)	année (58)
homme (21)	cru (20)	absorber (20)	légume (10)	saison des pluies
quelqu'un (18)	bleu (20)	boire (alcool) (8)	herbe (10)	eau (38)
être humain (7)	frais / froid (18)	sucer (10)	médicament (13)	ciel (35)
chose (5)	mouillé (10)	absorber (9)	plume (9)	pleuvoir (25)
enfant (5)	jaune (7)	abreuver (9)	aile (6)	dieu (11)
propriétaire (5)	humide (5)	se noyer (7)	tabac (5)	crachin (10)
AGENT (5)	herbe (5)	avaler (5)	livre (5)	Dieu (9)
tête (4)	nouveau (4)	manger (4)	branche (4)	nuage (7)
lieu (3)	gris (2)	soif (4)	balayer (4)	bruine (4)
parent (3)	perroquet (2)	bouche (4)	lettre (3)	rivière (3)
humain (2)	brut (2)	pleuvoir (3)	amulette (2)	foudre (3)
mâle (2)	doux (2)	téter (3)	palme (2)	arc-en-ciel (2)
corps (2)	violet (2)			

13	13	12	12	12
tout 'all'	voir 'to see'	(se) coucher 'to lie'	dormir 'to sleep'	mordre 'to bite'
121 sources	*231 sources*	*258 sources*	*308 sources*	*206 sources*
chaque (24)	regarder (79)	dormir (107)	couver (96)	piquer (35)
complètement (17)	trouver (47)	couver (11)	être couché (36)	manger (23)
tous (16)	retrouver (10)	tomber (10)	se coucher (25)	piquer (11)
entier (12)	obtenir (8)	lit (9)	coucher (21)	ronger (9)
immédiatement (9)	faire attention (8)	poser (9)	sommeil (20)	éteindre (6)
complet (8)	examiner (7)	habiter (7)	couché (18)	couper (6)
maintenant (6)	savoir (6)	étendre (5)	habiter (9)	dent (5)
bientôt (6)	rencontrer (5)	rester (4)	passer la nuit (8)	picorer (5)
partout (6)	visiter (4)	entrer (3)	s'allonger (7)	pincer (4)
jamais (5)	avoir (4)	passer la nuit (3)	se reposer (5)	croquer (4)
chacun (5)	apercevoir (3)	pencher (3)	rester (5)	arriver (3)
beaucoup (5)	comprendre (3)	se reposer (3)	copuler (4)	mâcher (3)
déjà (4)	gagner (3)			

12	12	11	11	11
s'asseoir, assis 'to sit'	sec 'dry'	être debout 'to stand'	femme 'woman'	langue 'tongue'
302 sources	*167 sources*	*179 sources*	*285 sources*	*151 sources*
habiter (113)	dur (33)	arrêter (78)	épouse (133)	parole (42)
rester (74)	sécher (27)	s'arrêter (39)	esclave (39)	langage (34)
habiter (14)	vide (12)	se lever (17)	femelle (36)	bouche (18)
être (14)	assécher (12)	attendre (8)	fille (10)	voix (11)
vivre (9)	tarir (7)	haut (6)	mère (8)	mot (9)
demeurer (7)	mince (5)	rester (5)	jeune fille (6)	palabre (8)
assis (6)	sécheresse (4)	cesser (4)	grand-mère (6)	parler (8)
poser (3)	séché (3)	s'élever (3)	soeur (4)	gorge (7)
devenir (3)	fané (3)	poser (3)	gauche (3)	discours (4)
attendre (2)	mûr (3)	construire (3)	parent (3)	conte (4)
mettre (2)	nu (3)	monter (3)	belle-soeur (3)	dire (3)

siège (2)	difficile (2)

11	11	11	11	11
long 'long'	marcher 'to walk'	noir 'black'	nuit 'night'	route , chemin 'road, path'
224 sources	*288 sources*	*262 sources*	*138 sources*	*271 sources*
grand (98)	aller (88)	sombre (170)	jour (17)	chemin (67)
loin (44)	partir (48)	sale (15)	obscurité (15)	fois (7)
haut (38)	voyager (32)	noircir (14)	noir (11)	voyage (7)
haute taille (11)	se promener (29)	bleu (12)	dormir (11)	sentier (6)
profond (10)	piétiner (15)	obscurité (12)	lune (5)	endroit (5)
hauteur (10)	bouger (9)	nuit (10)	soleil (5)	piste (4)
lointain (9)	venir (8)	bleu foncé (4)	mois (4)	lieu (4)
gros (5)	voyage (8)	Noir (3)	soir (4)	méthode (4)
large (5)	étendre (4)	obscur (3)	habiter (3)	voie (4)
être grand (4)	suivre (3)	foncé (3)	hier (3)	passage (4)
être haut (3)	courir (3)	rouge (2)	sombre (3)	rue (3)

10	10	10	9	9
blanc 'white'	feu 'fire'	montagne, colline 'mountain, hill'	coeur 'heart'	corne 'horn'
153 sources	*323 sources*	*330 sources*	*187 sources*	*122 sources*
Blanc (32)	enfer (41)	colline (114)	âme (50)	antenne (13)
clair (15)	chaud (32)	pierre (35)	esprit (49)	défense (13)
propre (15)	chaleur (27)	rocher (11)	foie (45)	bois (11)
briller (10)	lumière (18)	tas (8)	poitrine (16)	trompette (6)
Européen (8)	bois à brûler (13)	bosse (6)	estomac (5)	sifflet (5)
être clair (6)	fusil (10)	termitière (6)	nom (5)	trompe (4)
albinos (4)	chauffer (7)	pente (6)	vie (4)	fétiche (2)
brillant (3)	brûler (6)	sur (4)	intérieur (4)	flûte (2)
être propre (3)	chaud (être)	île (3)	conscience (3)	ventouse (2)
pur (3)	foyer (3)	dune (3)		

9	9	9	9	9
dent 'tooth'	eau 'water'	mourir 'to die'	ongle 'fingernail'	os 'bone'
184 sources	*235 sources*	*270 sources*	*201 sources*	*142 sources*
nom (37)	année (41)	tuer ? (102)	griffe (174)	arête (11)
trou + fossé (11)	pluie (38)	mort (40)	doigt (9)	noyau (11)
défense (8)	rivière (32)	finir (8)	sabot (7)	tibia (10)
ivoire (6)	liquide (9)	tuer (6)	écorce (5)	graine (6)
canine (5)	fleuve (5)	faner (5)	serre (5)	mourir (5)
mordre (5)	marée (4)	disparaître (4)	coquill(ag)e (5)	jambe (3)
bec (4)	lac (4)	tomber (3)	écaille (3)	squelette (3)
molaire (3)	mer (3)	sécher (3)	orteil (2)	côte (3)
pointe (2)	marigot (3)	se perdre (3)	peau (2)	genou (2)

9	9	8	8	8
rouge 'red'	viande 'meat'	entendre 'to hear'	nez 'nose'	pierre 'stone'
145 sources	*301 sources*	*294 sources*	*115 sources*	*164 sources*
mûr (36)	animal (168)	écouter (94)	morve (10)	rocher (40)
être mûr (20)	chair (61)	comprendre (89)	vie (9)	montagne (27)
jaune (15)	gibier (10)	sentir (63)	narine (4)	pierre du foyer (23)
mûrir (11)	animal sauvage (5)	creuser (27)	âme (3)	caillou (14)
chaud (5)	poisson (4)	percevoir (9)	front (3)	enclume (8)
marron (5)	sauce (4)	goûter (8)	germe (3)	colline (8)
orange (4)	muscle (4)	oreille (5)	bout (3)	meule (4)
cuit (3)	peau (3)	obéir (4)	extrémité (3)	trois (2)
brun (2)	corps (2)			

8	8	8	8	8
plein 'full'	quoi ? 'what?'	rond 'round'	savoir 'to know'	venir 'to come'
221 sources	*100 sources*	*81 sources*	*208 sources*	*296 sources*
remplir (63)	qui ? (16)	arrondir (16)	connaître (158)	arriver (48)
nombreux (29)	quel ? (13)	cercle (8)	comprendre (23)	aller (29)
beaucoup (27)	pourquoi ? (8)	tourner (4)	reconnaître (11)	sortir (18)
enfler (9)	comment (8)	tourner en rond (4)	pouvoir (9)	partir (12)
rassasié (6)	comment ? (7)	complet (3)	science (6)	marcher (8)
gonfler (6)	chose (6)	entier (3)	voir (6)	devenir (6)
rempli (4)	où ? (4)	circulaire (3)	connaissance (6)	apparaître (5)
complet (3)	lequel (3)	courbé (2)	sentir (2)	retourner (5)

7	7	7	7	7
cheveu/x 'hair'	oeil 'eye'	plume 'feather'	sable 'sand'	sang 'blood'
215 sources	*179 sources*	*150 sources*	*99 sources*	*105 sources*
poil (72)	visage (94)	poil (71)	poussière (36)	huile (6)
plume (36)	surface (38)	cheveu (19)	terre (20)	sève (5)
cheveu, poil (12)	graine (13)	aile (18)	sol (11)	veine (5)
feuille (8)	couleur (5)	cheveux (17)	grain (4)	rosée (4)
serpent (5)	devant (5)	fourrure (13)	île (3)	saigner (4)
tête (3)	jalousie (3)	feuille (9)	plage (2)	vin (3)
crinière (3)	face (3)	pelage (3)	lande (2)	pus (2)

6	6	6	6	6
cou 'neck'	graisse 'grease'	lune 'moon'	nom 'name'	nuage 'cloud'
167 sources	*160 sources*	*363 sources*	*193 sources*	*92 sources*
gorge (85)	huile (98)	mois (295)	dent (37)	brouillard (27)
voix (29)	gras (37)	clair de lune (42)	mot (14)	ciel (22)
nuque (14)	huile de palme (17)	nuit (5)	réputation (9)	pluie (7)
langue (langage) (6)	huile sp. (4)	menstruation (4)	appeler (6)	brume (4)
parole (3)	moelle (4)	saison (4)	coeur (5)	rosée (2)
col (3)	beurre (2)	miroir (3)	célébrité (3)	fumée (2)

6	6	6	5	5
oreille 'ear'	voler 'to fly'	un 'one'	cendre/s 'ash/es'	nager 'to swim'
129 sources	*155 sources*	*282 sources*	*114 sources*	*89 sources*
feuille (31)	sauter (53)	quelqu'un (49)	poussière (49)	se baigner (3)
bord (6)	se lever (35)	seul (24)	farine (3)	ramer (3)
entendre (5)	s'envoler (5)	même (19)	charbon (2)	laver (3)
lobe (3)	courir (3)	un seul (13)	poudre (2)	couper (3)
lobe de l'oreille (3)	plume (3)	le même (9)	être blanc (2)	flotter (3)
écouter (2)	flotter (2)	unique (5)		

5	5	4	4	4
nouveau 'new'	sein 'breast'	foie 'liver'	fumée 'smoke'	je 'I'
212 sources	*174 sources*	*103 sources*	*76 sources*	*126 sources*
jeune (112)	lait (75)	coeur (46)	vapeur (17)	3sg (16)
frais (10)	poitrine (26)	herbe (5)	poussière (6)	2sg (12)
récemment (3)	pis (18)	colère (5)	tabac (5)	même (9)
vert (3)	mamelle (13)	esprit (3)	nuage (3)	1pl (3)
cru (3)	dans (3)			

4	4	4	4	4
nous 'we'	oiseau 'bird'	poisson 'fish'	queue 'tail'	qui ? 'who?'
110 sources	*179 sources*	*131 sources*	*86 sources*	*126 sources*
2pl (8)	animal (7)	poisson (71)	bout (3)	quoi ? (15)
1sg (3)	poulet (6)	pêcher (6)	racine (2)	quel ? (6)
3sg OBJ (3)	pintade (4)	viande (4)	fin (2)	propriétaire (4)
2sg POS (3)	chauve-souris (2)	animal (3)	crin (2)	que (4)

4	4	3	2	2
tu 'you (sg)'	vous 'you (pl)'	genou 'knee'	deux 'two'	oeuf 'egg'
139 sources	*132 sources*	*57 sources*	*91 sources*	*111 sources*
2pl (26)	3pl (19)	coude (3)	sept (5)	pondre (4)
3sg (12)	2sg (13)	jarret (2)	paire (2)	testicule (3)
1sg (13)	3sg (11)	noeud (2)		
1pl (4)	1pl (7)			

2	1	1
pou 'louse'	chien 'dog'	étoile 'star'
117 sources	*74 sources*	*49 sources*
puce (9)	pauvre (8)	luciole (6)
épouiller (2)		