



# How a West African language becomes North African, and vice versa

Lameen Souag

## ► To cite this version:

Lameen Souag. How a West African language becomes North African, and vice versa. *Linguistic Typology*, 2021, 10.1515/lingty-2021-2083 . halshs-03507171

**HAL Id: halshs-03507171**

**<https://shs.hal.science/halshs-03507171>**

Submitted on 3 Jan 2022

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

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

---

**Erratum:** Denis Creissels assures me that Bambara *jíri* can mean “tree” and “wood”, but never “firewood”; please correct accordingly.

---

## How a West African language becomes North African, and vice versa

Lameen Souag

Updating the methodology of Hayward (1991) using the concept of colexification (François 2008), this paper, for the first time, provides quantitative evidence that the languages of the West African Sahel/Savanna form a lexical-typological language area characterised by shared colexifications absent further north. It then uses the comparative method to determine how languages entering or leaving this area, or coming into increasing contact with it at its edges, have converged with their new neighbours within the past millennium. The results indicate sharp differences in the respective roles and rates of borrowing and calquing, with the latter acting almost exclusively to increase shared colexifications.

### 1 Introduction

Cross-linguistic typological variation shows a strong tendency towards areal distribution, with neighbours more likely to share properties with each other than with languages further afield. For grammatical typology, this applies not only within relatively limited regions (Muysken 2008), but even at continental scales (Dryer 1989). Insofar as it applies across mutually unrelated languages, this fact can only be explained as a result of long-term language contact.

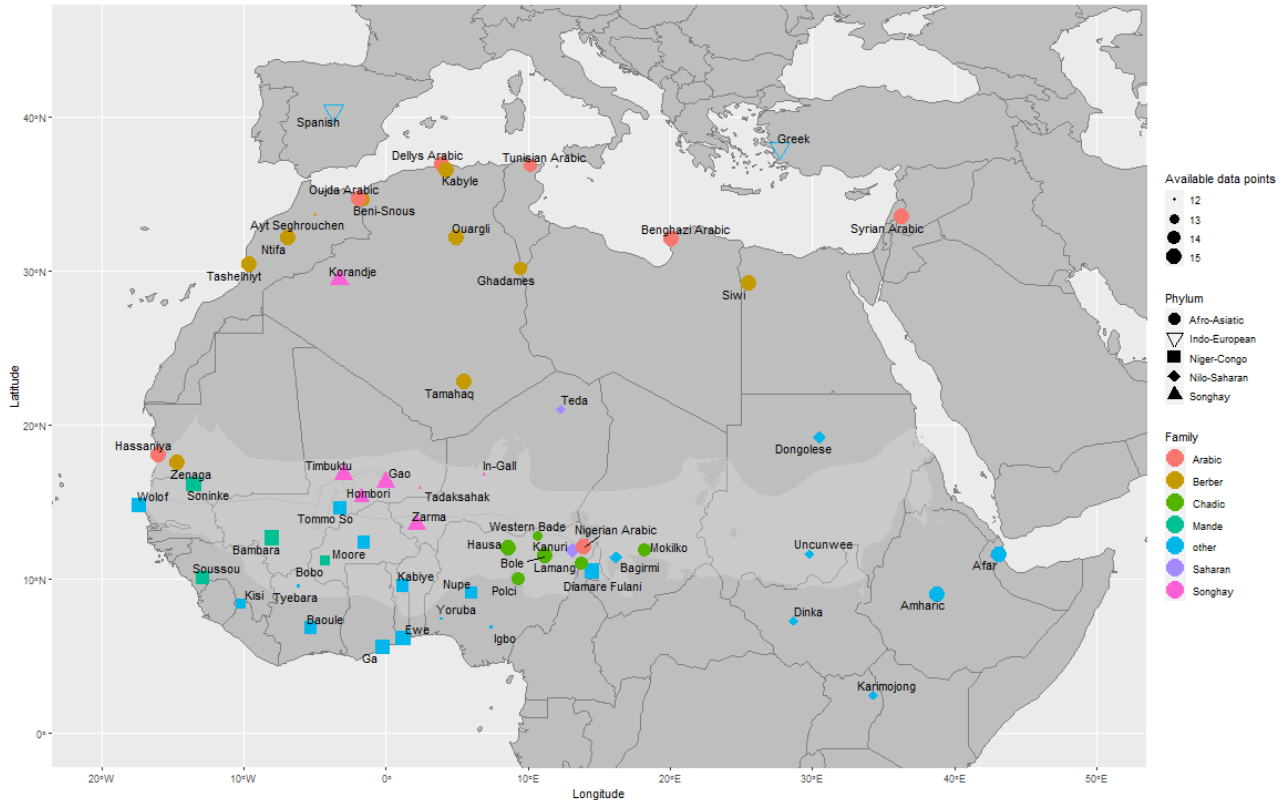
An increasing body of work suggests that the same is true for lexical typology (Koptjevskaja-Tamm & Liljegren 2017; Schapper, San Roque & Hendery 2016). The extent of lexical-typological convergence between different languages in a linguistic area can be quantified by counting shared lexical-typological features, a methodology pioneered by Hayward (1991; 2000) and Smith Stark (1994) and formalized here (Section 2) using the concept of colexifications (François 2008). But to move beyond description towards explanation, it is necessary to understand how exactly this convergence happens. Which specific diachronic changes are involved? Which types of change are more or less important? How quickly do such changes proceed? In situ contact provides some useful examples (Sections 3.1, 3.2), but in the absence of adequate written historical evidence, such changes can most effectively be identified in cases where an offshoot of a language group spoken mainly in one area moves into a different one not sharing the same properties.

The linguistic areal boundary between North Africa and the West African savanna belt, stretching through the Sahel – shown here for the first time to be a lexical-typological boundary as well (Section 3) – provides us with two such cases, similar in age but very different in outcome, where individual colexifications’ histories can be reconstructed case by case (Section 4). The Northern Songhay languages, a branch of Songhay which has moved north into the Sahara mostly within the past 700 years, demonstrate the possibility of relatively rapid lexical-typological convergence by the loss of colexifications, as is most clearly manifested in its northernmost member, Korandje. Nigerian Arabic, on the other hand – which reached the Lake Chad region from the east starting about the same period – exhibits much slower convergence through gaining colexifications. All but one of the changes observed can straightforwardly be categorized as either borrowing or calquing,

and closer examination permits conclusions to be drawn on the relative roles of these two processes (Section 5).

## 1.1 Areal context

*Map 1: Languages examined*



North Africa and West Africa show strikingly different patterns of language distribution. North Africa and the Sahara form a classic spread zone (Nichols 1992; Souag 2017), dominated by the low-diversity expansions of Berber and Arabic (both belonging to the Afroasiatic phylum). The semi-arid lands just south of the Sahara – traditionally called the Sudan, but labelled Sahel/Savanna here to prevent confusion with the state of that name – are characterised by a far higher degree of linguistic diversity in genealogical terms, involving at least three phyla (Niger-Congo, Nilo-Saharan, and Afroasiatic) even by the rather inclusive standards of Greenberg (1963). Among these, a few stand out as particularly widespread precolonial lingua francas, from west to east: Bambara (Mande, Niger-Congo) in the upper Niger, Songhay (isolate/Nilo-Saharan) in the Middle Niger, Hausa (Chadic, Afroasiatic) between the Niger and Yobe rivers, Kanuri (Saharan, Nilo-Saharan) around Lake Chad, and Fulani (North Atlantic, Niger-Congo) scattered throughout the area. For a full list of languages examined, along with corresponding references and classifications, see Table 8 in Supplementary materials.

Environmentally speaking, the Sahel/Savanna consists mainly of two ecoregions, marked in light grey in Map 1 above: the Sahelian acacia savanna stretching from northern Senegal in the west to central Sudan in the east, and the West Sudanian savanna to its south (Olson et al. 2001). Contact between north and south has been significant, particularly following the introduction of the camel and the rise of the trans-Saharan caravan trade (Fentress 2009). Nevertheless, the relative difficulty

of crossing the Sahara between north and south contrasts with the much greater ease of travelling east and west on either side of it. This applies particularly to lowland regions, with the highlands of eastern Chad slowing down travel between the western and eastern Sahel/Savanna. This suggests that linguistic features should diffuse faster within the western Sahel/Savanna than across the Sahara, in which case the western Sahel/Savanna should be expected to show some characteristics of a linguistic area.

Among existing areally focused work on typological features in North Africa and the West African Sahel/Savanna (Caron & Zima 2006; Chaker & Caubet 1996), some indications already emerge that the southern boundary of Arabic and Tuareg – roughly corresponding to the southern boundary of the geographical Sahel – is also an areal boundary. In particular, Clements and Rialland (2008) identify North Africa as the core of a phonological area including the Sahara, and the Sudanic region (including the Sahel) as another distinct phonological area. In the areal model of Güldemann (2018), combining phonological and grammatical features, North Africa and the Sahara are classed with Egypt and the Middle East in the Afroasiatic spread zone, whereas the Sahel is treated as a transition zone towards the Macro-Sudan belt to its south. However, lexical typological features have hardly been explored cross-linguistically across these regions, except in the context of possible pan-African features (Heine & Leyew 2007: 24–26). It will be shown here that, while few lexical-typological features seem to unite the Sahel/Savanna as a whole, the major inter-ethnic languages of the western Sahel/Savanna – Bambara, Songhay/Zarma, and Hausa – show striking commonalities in colexification patterns despite being unrelated. The number of these features found in neighbouring languages diminishes as one goes north or south from the core Western Sahel/Savanna region. Insofar as these lexical-typological commonalities are created and maintained by mutual contact determined by geographical factors, one would expect languages moving into and out of this zone to change their lexical structure accordingly. This expectation is borne out, as will be seen below (Section 4).

## **1.2 Trans-areal languages**

In the context of a potential areal boundary between North Africa and the Sahel/Savanna, two of the languages examined in this article stand out for having crossed this boundary within the past millennium. These two trans-areal languages are particularly useful for revealing the mechanisms underlying areal convergence here, and will thus be given more detailed treatment.

Korandje (Souag 2010; Tilmatine 1996; Cancel 1908), spoken at the small oasis of Tabelbala in southwestern Algeria, belongs to the Songhay family (Nilo-Saharan according to Greenberg (1963), though the evidence for Songhay being Nilo-Saharan is weak and disputed). As such, it stands out as an outlier within the northern Sahara, an area dominated throughout its recorded history by just two linguistic expansions: first Zenati Berber, then Maghrebi Arabic (both Afroasiatic). The Songhay family originated more than a thousand kilometres to the south, and is typologically more closely aligned with other languages of the western Sahel/Savanna than with Berber or Arabic. Songhay first reached Tabelbala sometime around the thirteenth century (Souag 2015a), giving it ample time to come under the influence of regional lingua francas. Ever since about 1600, Tabelbala has been politically and culturally dominated by groups from the north, and multilingualism has been a valuable asset; for the past century at least, almost all Korandje speakers have been bilingual in Maghrebi Arabic, and the distribution of borrowings suggest that they had previously often been bilingual in Berber. For its lexicon, I rely on my own field data.

Nigerian Arabic is a western Sudanic Arabic variety at the extreme end of a dialect continuum stretching across Chad to Darfur, traditionally spoken by nomadic cattleherds who started reaching the Lake Chad area in the 14<sup>th</sup> century (Owens & Hassan 2008). Despite some encroachment by Hausa (Chadic, Afroasiatic), the regional lingua franca was and largely remains Kanuri (Saharan, arguably Nilo-Saharan), coexisting with a number of smaller Chadic languages. Kanuri influence, strengthened by the increasing urbanization of Nigerian Arabic speakers which brings them into more regular daily contact with the more numerous Kanuri speakers, is manifest not only in lexical borrowings but also to a striking extent in lexical typology, including the emergence of a major word class of ideophones (Owens & Hassan 2004) and the calquing of idioms and associated polysemy (Owens 2015). Its lexicon has been documented principally by Kaye (1982; 1987).

## 2 Methods

In order to reach any conclusions on the effects of language contact on lexical typology in North and West Africa, it is necessary to consider how lexical typology can be compared across languages without falling foul of the linguistic relativity of emic concepts (2.1) and how to make such comparisons quantitative (2.2). We then select a list of specific variables to compare (2.3), including appropriate phrasal contexts. Examining the worldwide distribution of these variables makes it possible to provide a preliminary estimate of expected sharing across languages in the absence of contact and inheritance effects, providing a null hypothesis to compare to the actually attested results outlined in Section 3. Finally, summary information is provided on the languages to be examined (2.4 and Supplementary materials).

### 2.1 Definitions

Whereas considerable work has been done on how to quantify the extent of lexical borrowing in a language usefully (e.g. Haspelmath & Tadmor 2009), no generally accepted methods exist for quantifying the effects of lexical-typological restructuring. The quantification of lexical-typological features to identify language areas seems to have first been explored for the Ethiopian area, in pioneering work by Hayward (1991; 2000), and on a larger scale for Mesoamerica (Smith Stark 1994). Hayward's approach is readily generalisable: he sets up a list of lexicalization patterns characteristic of the Ethiopian area – cases “where a single word embraces two very distinct senses or where the structure of a word indicates a particular derivational formation” – and counts how many of them are shared between given pairs of languages. The main problem with this approach is the question of whether two given meanings count as distinct senses or not. This issue can be avoided by instead counting *colexifications*, adopting the terminology of François (2008: 170) who defines the term as follows:

A given language is said to colexify two functionally distinct senses if, and only if, it can associate them with the same lexical form.

In this definition, François' unfortunately ambiguous term “functionally distinct senses” is explicitly intended to refer to any potentially differentiable meanings, whether the speakers encode them as psychologically distinct senses, contextual readings of the same sense, or even accidental homophones (the difference between those three possibilities being strictly emic and as such internal to a specific language):

For example, the fact that French lexifies <rectilinear> with <honest> but not with <frank> suffices to justify the choice of distinguishing between the two latter meanings as if they were two separate senses. Even though this may fail to represent faithfully the language-internal perception of an English native speaker, at least this serves efficiently the purpose of cross-linguistic comparison: it becomes then easy to state the facts by saying that these two senses are treated the same in English, and not in French. (*ibid*:168-169)

In fact, any semantic distinction lexicalised in any language can be incorporated into an etic grid of meanings – understood as grounded in extra-linguistic contexts of usage, not as definitions or maximally economical concepts – usable for stating a colexification, irrespective of its emic status in any other language. Japanese has distinct words for ‘cool water’ (*mizu*) vs. ‘hot water’ (*yu*); we can say that English colexifies *mizu* and *yu* without any (absurd) suggestion that English speakers conceptualise the word *water* as having two distinct senses depending on temperature. In principle, we need not even require that the distinction be lexicalised in some language; it is probably true, and certainly meaningful, to state that all languages colexify ‘cool water’ with ‘cold water’. For comparative purposes, however, it is sufficient that the etic grid should consist of contexts of usage sufficiently well-defined to be clearly translatable into any of the languages involved. All colexifications used in this paper are accordingly defined not just through glosses but through short phrases narrowing down the intended context of usage of those glosses (as listed in 2.3). A non-verbal stimulus-based approach, which would in principle be ideal for avoiding translation effects in the elicitation of etic distinctions, was not practicable in this case due to the number of languages examined.

Unlike polysemy (in which some approaches allow one sense to be considered as psychologically primary and others as derivative), or grammaticalisation (in which one sense is taken as historically original and others as derivative), colexification is strictly symmetrical and commutative: if  $X = Y$ , then  $Y = X$ . The order of the two meanings is purely conventional.

Using the notion of colexification, we can then examine shared colexifications and shared non-colexifications, either of which can plausibly result from mutual contact, and contrast them with non-shared colexifications, which are unlikely to reflect mutual contact. For example, Korandje *kadda* is used for both ‘small (in size)’, as in *aḍṛa kadda* “small mountain (hill)”, and ‘young (in age)’, as in *ʕ-ba-təndz kadda* “I’m still (too) young (to retire)”. Algerian Arabic *ɣiɾ* is used in the same two contexts: one can say *ħajra ɣiɾ-a* “a small rock”, with no possible implication of recency, or tell even a tall man *mazal-ək ɣiɾ* “you’re still young”. Korandje and Algerian Arabic thus *share the colexification* of ‘small (in size)’ with ‘young (in age)’. On the other hand, with a mass noun, Korandje *kadda* is also used for ‘a little’, as in *ir kadda* “a little water”; Algerian Arabic does not use *ɣiɾ* in this context, instead using *šwiyya* (*šwiyya ma* “a little water”). Algerian Arabic and Korandje thus *do not share the colexification* of ‘small’ with ‘a little’. Kabyle Berber, however – like Algerian Arabic – does not colexify ‘small (in size)’ (*aməštuh*) and ‘a little’ (*šituh*); it follows that Algerian Arabic and Kabyle *share the non-colexification* of these two senses, i.e. they agree on not colexifying them.

## 2.2 Measurement

In lexical comparison, the proportion of shared lexicon is normally estimated by counting cognate items on a fixed list of basic meanings translatable into all the languages involved. To compare polysemy, one might adopt an analogous approach using a list of pairs of meanings rather than

individual meanings. If two languages both colexify the pair, or both fail to colexify it, then they will be considered to agree on that pair; if one colexifies it and the other does not, then they disagree. For a given list of meaning pairs, the similarity score  $s_{ij}$  between two languages ( $i, j$ ) is the number of meaning pairs on which they agree ( $n_{ij}$ ), divided by the total number of meaning pairs included in the list ( $N$ ):  $s_{ij} = n_{ij} / N$ .

However, unlike a wordlist, a list of colexifications also allows for a single-language measure :  $s_i = n_i / N$ , where  $n_i$  is the number of colexifications in the list that are present in language  $i$ . If the list consists of a representative sample of “natural” colexifications with a worldwide distribution, then this measure might be taken as representing the language’s propensity for colexification in general. If it consists of ones with a specifically areal distribution, as here, then it may be taken as a rough measure of the degree to which the language belongs, in terms of its lexical typology, to the language area in question.

### 2.3 List selection

Obviously, such measures are highly sensitive to the selection of meaning pairs. In lexical comparison, the best-known word lists selected for comparison focus on basic vocabulary – meanings supposed to be familiar in virtually every culture, as in the Swadesh list (Swadesh 1952) – or on the least commonly borrowed vocabulary, as in the Leipzig-Jakarta list (Tadmor 2009). By analogy, one might pick a list of meaning pairs for which colexification appears to be particularly natural, in the sense of recurring independently all over the world. Such a list, unlike the Swadesh or Leipzig-Jakarta lists, would be appropriate for a worldwide comparison of polysemy patterns.

However, the goal at hand is to examine areal contact effects in a particular region, not to maximise cross-linguistic comparability. For this purpose, we need a list selected to maximise the contrast between north African languages and mainstream Songhay varieties, while at the same time restricting itself to meanings lexically expressed in all three and (as far as possible) ascertainable using published materials. In a first pass, such a list was selected by going through Heath’s (1998) dictionary of Gao Songhay (Koyraboro Senni) and selecting words with two senses, both reasonably basic, that would not be colexified in Algerian Arabic, chosen for convenience as a representative North African variety for whose lexicon the author would not be forced to rely exclusively on published lexical data. Since Arabic and Berber systematically differentiate verbs from nouns through their form, making strict colexification between them impossible, only senses sharing the same part of speech were chosen, ruling out cases like Djenne Songhay *kamba* = ‘hold onto (v.) = hand (n.)’. (In Songhay, Berber, and Arabic, “up/above” was analysable as a noun, making colexification with “sky” possible.) In subsequent passes, colexifications found to be widespread in both Songhay and North Africa (e.g. ‘ripen = cook’, ‘do = put’, ‘stand = stop’, ‘stand up = wake up’) were removed in order to maximise contrast. It also proved necessary to remove colexifications involving words that were absent from too many of the available dictionaries (eg Sahelian ‘youngest child = little finger’, North African ‘wife’s brother = wife’s father’, ‘smile = laugh’). Finally, the worldwide distribution of each colexification was examined, based primarily on CLICS<sup>2</sup> (List et al. 2018) and Heine and Kuteva (2002: 84–86), and supplemented from other sources where available. This list is constructed to maximise the contrast between Songhay and North Africa, and cannot thus be used to argue for the existence of such a contrast. However, the manner of its construction does not predetermine the results for other languages of the Sahel/Savanna area; rather, it makes it possible to see whether these languages behave more like Songhay or like North Africa.

This yielded the following list of fifteen colexifications characteristic of Songhay – and, as it turned out, of the Sahel – but not of North Africa, ordered below in descending frequency (given in brackets) within CLICS<sup>2</sup> (List et al. 2018) to get an idea of their worldwide distribution out of the 1220 varieties covered there to varying degrees, and accompanied where possible by indications of the relative frequency of one sense as a colexification of the other. The phrases given between brackets for each colexification are provided to narrow down the intended meaning through collocation or definition, removing as far as possible the ambiguity inherent in English glosses.

1. ‘with (comitative) = with (instrumental)’, e.g. “travel with a friend” vs. “cut with a knife”, for which cf. Stolz et al. (2011); out of 322 languages examined there, only 76 – less than a quarter – unambiguously show this colexification, with a particular concentration in western Eurasia and western sub-Saharan Africa. Heine and Kuteva (2002: 84–86) take the directionality to be COMITATIVE > INSTRUMENTAL, while noting that “more diachronic evidence is required”.
2. ‘hand (from fingertips to wrist) = arm (from wrist to shoulder)’, possibly correlated with clothing conventions and hence latitude (Brown 2013). I exclude cases where ‘hand’ is used as the cover term for ‘hand + arm’, but where a distinct word for ‘arm’ exists. (196, across much of the world; ‘arm’ and ‘hand’ are each other's most frequent colexification.)
3. ‘hurt = be ill’, e.g. “his head hurts” vs. “he is ill”). Within CLICS<sup>2</sup>, ‘painful = sick’ is attested 159 times in and near Australia, while ‘pain = sick’ is widespread in East Asia and the Caucasus (107 times; each other's most frequent colexification).
4. ‘easy = light (weight)’, e.g. “this task is easy” vs. “this rock is light” (103, largely in western Eurasia; each other's most frequent colexification).
5. ‘cover = close/shut’, e.g. “cover him with a blanket” vs. “shut the door” (43, mostly in inner/northern Eurasia; also common in Insular Southeast Asia [Antoinette Schapper, p.c.]; each other's most frequent colexification).
6. ‘top/above = sky’, e.g. “put it on top”, “the light is above the table” vs. “there are clouds in the sky”; cf. Heine and Kuteva (2002: 79–80), who consider TOP > UP “the primary source for UP markers” in “much of the southern half of Africa”, based on cases where the two meanings are colexified there. Within CLICS<sup>2</sup>, ‘top = sky’ appears 7 times, ‘sky = above’ 40 times (all but one within sub-Saharan Africa), and ‘sky = up’ 11 times, mainly in South America. The most frequent colexification of ‘above’ is “sky”, but not vice versa, and ‘sky’ is only the eighth most frequent colexification of ‘top’.
7. ‘tree = firewood’, e.g. “sit in the shade of this tree” vs. “she's carrying firewood”) (38, concentrated in inner Asia and South America; found in 18.5% of Papuan+Australian languages, cf. Schapper et al. (2016)). While ‘tree’ and ‘wood’ are each other's most frequent colexification, ‘firewood’ is only the fourth most frequent colexification of ‘tree’, and ‘tree’ the third of ‘firewood’.
8. ‘kill = extinguish/put out’, e.g. “kill the pig” vs. “put out the fire” (35, mainly sub-Saharan Africa, but also e.g. in the Caucasus and Polynesia; also common in Insular Southeast Asia [Antoinette Schapper, p.c.]). ‘Kill’ is the most frequent colexification of ‘extinguish’ (but not vice versa).



9. 'separate = divorce', e.g. "he separated them into groups" vs. "he divorced his wife", possibly correlated with differences in marriage conventions. (25, mainly Bantu). 'Separate' is the most frequent colexification of 'divorce' (but not vice versa).
10. 'father = father's brother', characteristic of Iroquois-type kinship systems (Morgan 1871; Murdock 1981; Liljegren this issue). (23, mainly South American). 'Father's brother' is the most frequent colexification of 'father' (but not vice versa).
11. 'hard (surface) = difficult', e.g. "this wood is hard" vs. "this task is difficult" (22). 'Hard' is the second most frequent colexification of 'difficult' (after 'heavy'), but not vice versa.
12. 'stranger/outsider = guest', e.g. "he doesn't trust strangers" vs. "he prepares good dinners for his guests" (16, concentrated mainly in northern Nigeria and premodern Europe; each other's most frequent colexification).
13. 'mouth = edge', e.g. "put food into one's mouth" vs. "edge of the road" (16). 'Edge' is only the fourth most frequent colexification of 'mouth', and 'mouth' the seventh of 'edge'. Not found in Heine and Kuteva (2002: 214), but some of their examples for MOUTH > FRONT may in fact reflect this colexification.
14. 'hear = smell (v.)', e.g. "I heard a sound" vs. "I smelt an odour" (14). 'Hear' is only the third most frequent colexification of 'smell', and 'smell' the seventh of 'hear'.
15. 'foot = time (instance, French *fois*)', e.g. "a person has two feet" vs. "he did it four times"; the relevant sense of 'time' is not in CLICS<sup>2</sup>, but the colexification is found outside the region in Biblical Hebrew *páʕam* (Brown, Driver & Briggs 1906: 822).

All of these colexifications are obviously motivated by nearly universal facts about human cognition and the environment we live in, and none of them is limited to a single part of the world. Nevertheless, each of them is individually absent from most languages; the most widely attested, 'with (comitative) = with (instrumental)', seems to occur in less than a quarter of the world's languages. The combined presence of many of them thus tells us much more than the presence of any one of them individually. This can be roughly estimated quantitatively. Apart from considerations of language contact and common origin, there is no obvious reason to assume that the presence of one of these colexifications makes another inherently more likely. Assuming that the probability of presence of each one is independent of the others, and that each one has a probability as high as one in five of occurring spontaneously (which, judging by CLICS<sup>2</sup>, is likely to be an overestimate), then most languages (65%) are expected to share less than four of these colexifications, and practically all languages (98%) less than seven. The probability of independently sharing 10 or more of these colexifications would be  $0.2^{15} + 15(0.2^{14})(0.8) + (15 \cdot 14/2!)(0.2^{13})(0.8^2) + \dots + (15 \cdot 14 \cdot \dots \cdot 10/10!)(0.2^{10})(0.8^5)$ , which comes to less than 0.00012 – i.e., one in 8831, more than the estimated number of languages on earth. In the absence of contact and common ancestry, we would then expect not to find any other language on earth sharing as many as ten of these colexifications with Gao Songhay, let alone one nearby. If we do find several such cases in the same geographical area, genetically unrelated, then that must be taken as a strong indicator of the importance of contact in causing these colexifications.

To give some idea of the combined frequency worldwide of these colexifications in practice outside of the area in question, English shares two of the 15 (*with*, *hard*); Burmese (San San Hnin Tun, p.c.)

shares four ('with (comitative) = with (instrumental)', 'hand = arm', 'kill = extinguish', 'separate = divorce'); Limbu in Nepal (Michailovsky (2002) and p.c.) shares six ('hand = arm', 'hurt = ill', 'tree = firewood', 'kill = extinguish', 'hard = difficult', 'hear = smell'); colloquial Indonesian (Antoinette Schapper, p.c.) shares seven ('with (comitative) = with (instrumental)', 'arm = hand', 'hurt = be ill', 'cover = close', 'kill = extinguish', 'separate = divorce', 'father = father's brother').

Ideally, it would be useful to have an equally long list of colexifications identified as characteristic of Northwest Africa but not the Sahel, made for example by listing words colexified in Algerian Arabic but not in Gao Songhay. In practice, this proved unfeasible. It was much easier to identify colexifications characteristic of the Sahel and rare or absent in North Africa than vice versa, despite the Sahel's considerably greater linguistic diversity; such cases of the latter as were found often involved senses too rarely documented in available dictionaries. Pending further research, it remains unclear whether this reflects a Sahelian preference for relying on context rather than stem choice to disambiguate semantically related senses, or arises simply because many of the senses colexified in North African languages are also colexified in English and French, making them invisible to most dictionary-writers.

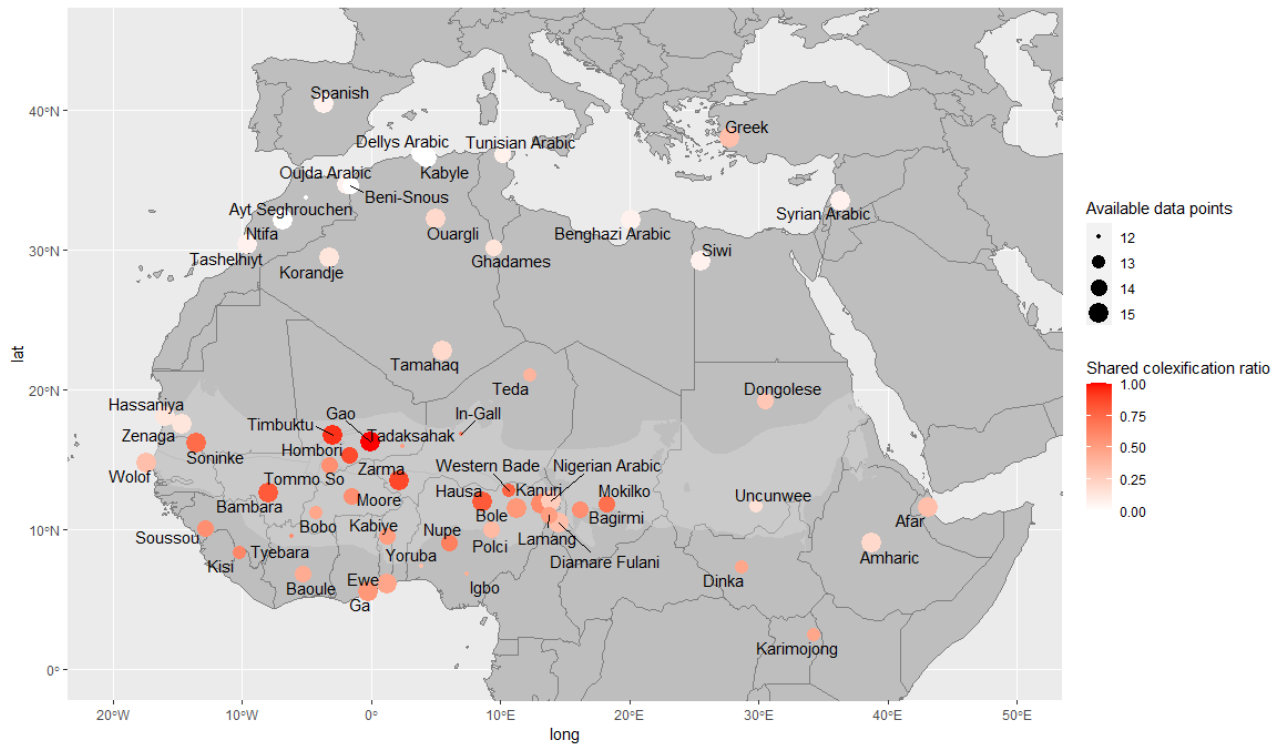
## **2.4 Language selection**

For this study, colexification lists were gathered for a total of 69 varieties in and near the region (see Supplementary materials for details and sources). If less than 12 of the pairs of the pairs of meanings listed below could be determined from the sources examined, no colexification ratio was calculated, and the variety was excluded from Map 2. This left 58 varieties with sufficiently complete materials to be used: 16 from North Africa (defined to include Zenaga and Hassaniya), 34 from West/Central Africa (26 of them in the Sahel/Savanna), 2 from Europe, 1 from the Middle East, and 5 from East Africa. Another five Songhay varieties, and two Arabic varieties, were used to facilitate comparative-historical reconstruction, while a total of 10 varieties with inadequate lists were used to elucidate the distribution of specific colexifications. Because of the particular diachronic interest of Songhay as a family straddling the Sahel-Sahara boundary, as many varieties of Songhay were examined as feasible.

## **3 The distribution of Sahelian colexifications**

Simply measuring the proportion in each language of the 15 colexifications defined in 2.3, as calculated by dividing the number of these colexifications present by the number of colexifications which could be checked using available data, draws a striking contrast not just between north Africa and Songhay (by construction), but between North Africa and the western Sahel/Savanna as a whole; see Map 2. A closer look reveals the effects of contact with West Africa on languages of the Sahara (3.1) and of contact with one another on languages of the western Sahel/Savanna (3.2).

*Map 2. Proportion of matches by language, produced using R (R Core Team 2018)*



### 3.1 In North Africa and the Sahara

In north Africa proper, north of the Atlas Mountains, the fifteen Sahelian colexifications are almost entirely absent, by construction, as indeed in other Arabic varieties examined. ‘Hand = arm’ (*yidd*) pops up in Tunisian Arabic (along with some Middle Eastern varieties, cf. Behnstedt and Woidich (2010: 143)), and ‘hurt = be ill’ (*aḡn*) in Tashelhiyt Berber, but none of the others are attested there. In Saharan Berber varieties, on the other hand – especially north and central Saharan ones – the picture is a little different, as illustrated by Table 1.

Table 1. Sahelian colexifications across Arabic and Berber varieties of the Sahara.

|                             | Zenaga         | Ouargli       | Ghadames      | Tamahaq          | Siwi       |
|-----------------------------|----------------|---------------|---------------|------------------|------------|
| ‘kill = extinguish’         | -              | <i>nəγ</i>    | <i>ānn</i>    | -                | <i>nay</i> |
| ‘outsider = guest’          | <i>iɁmāški</i> | -             | <i>anazil</i> | <i>amağar</i>    | -          |
| ‘tree = firewood’           | <i>āššaɁr</i>  | -             | -             | -                | -          |
| ‘divorce = separate’        | -              | <i>əbɁa</i>   | -             | <i>māzzāy</i>    | -          |
| ‘father = father's brother’ | -              | -             | -             | <i>ti-, abba</i> | -          |
| ‘top = sky’                 | -              | <i>ažanna</i> | -             | -                | -          |

Each of these colexifications is shared with Bambara, (Gao) Songhay, Hausa, and Kanuri alike, except ‘divorce = separate’ (missing from Bambara) and ‘tree = firewood’ (missing from Kanuri). Contact with those four languages in earlier centuries have been significant throughout the region, these being at once the languages most useful to southbound traders and the lingua francas

most likely to be known by enslaved people; loans from all four are well attested across Saharan Arabic and Berber varieties (Souag 2013). The appearance of these colexifications in Saharan Berber varieties thus seems likely to be a result of their influence in some cases. For example, Proto-Berber ‘extinguish’ seems to be *\*s-əysəy* (the causative of *\*ăysəy* ‘be extinguished’), contrasting with ‘kill’ *\*ănyəʔ*; cp. Kabyle *ss-əxsi* vs. *nəy*, Tamahaq *s-ăysəy* vs. *əny*, and – in the same Zenati branch as Ouargli – Figuig *ss-əxsəy* vs. *nəy* (Kossmann 1997: 519). There is no obvious Berber-internal motivation for the loss of *\*s-əysəy*, and L2 speakers whose native language colexifies ‘extinguish = kill’ would be likely to use reflexes of *\*ănyəʔ* to express both (see section 4).

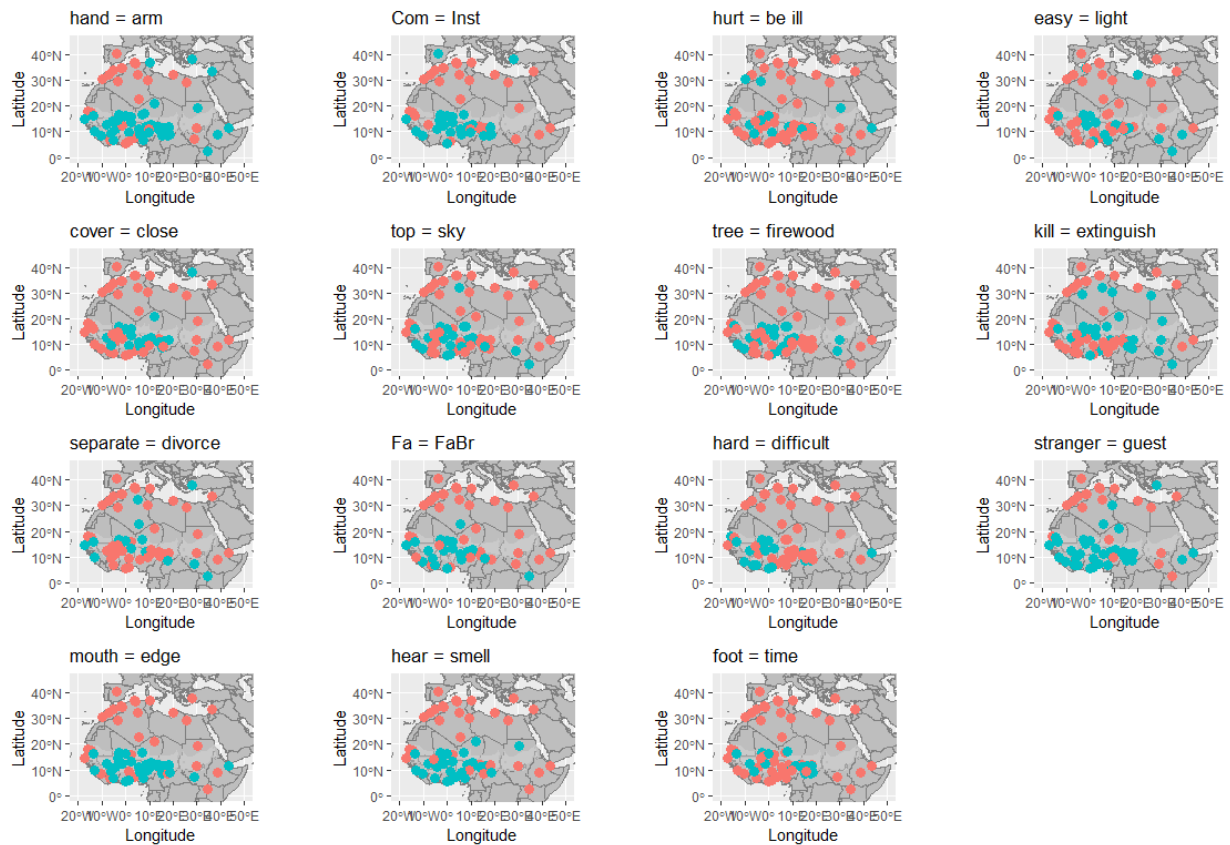
On the other hand, the general absence of some of these colexifications may in some cases reflect influence from Arabic. There is no reconstructible proto-Berber term for ‘father's brother’ as distinct from ‘father’; all northern Berber varieties seem to express the concept with the Arabic borrowing *ʕammi*. It is plausible in this case that the distinction was introduced to Berber as part of a wholesale restructuring of kinship terminology to match an Arabic model.

The only Saharan Arabic variety examined, Hassaniya (Mauritania), has *uʒaʃ* ‘hurt = be ill’ and *gāsi* ‘hard = difficult’. Both colexifications are shared with Bambara, and the latter at least with Soninke, both of which are known to have been relevant contact languages in the recent past. Neither is found in Wolof.

### 3.2 In the Sahel

Many of the colexifications examined are pervasive throughout sub-Saharan West Africa, including ‘with’, ‘hand = arm’, ‘stranger = guest’ (almost exceptionlessly), ‘father = father's brother’, ‘mouth = edge’, ‘hear = smell’. Others appear characteristic of the western Sahel/Savanna as opposed to the forest belt, notably ‘foot = time’, ‘cover = close’, ‘separate = divorce’, ‘kill = extinguish’ (see Map 3). Adding them all up together with the ones showing a less clear-cut distribution by latitude, we get a finer level of discrimination, as seen above in Map 2.

*Map 3: The distribution of individual colexifications (green = colexified, red = not)*



The list was constructed around Gao Songhay, which of course has the highest scores (see below). Songhay's neighbours Bambara (upstream along the Niger) and Hausa (to the east) likewise have particularly high scores (12 out of 15), along with the formerly more widespread regional language Soninke (11/15). Most of these shared colexifications can be explained neither by borrowing of forms nor by common ancestry, and can thus be explained only through calquing; Mande, Songhay, and Chadic belong to three different macrophyla, and even between the two Mande languages Soninke and Bambara, the words in question are usually not cognate, as seen in Table 2.

Table 2. Colexifications shared across languages of the Sahel on or near the Niger River.

|                      | Soninke      | Bambara       | Gao Songhay  | Hausa                              |
|----------------------|--------------|---------------|--------------|------------------------------------|
| ‘Com = Inst’         | -            | -             | <i>nda</i>   | <i>dà</i>                          |
| ‘hand = arm’         | <i>kíttè</i> | <i>bólo</i>   | <i>kabe</i>  | <i>hannū</i>                       |
| ‘hurt = be ill’      | -            | <i>bàna</i>   | <i>dor</i>   | <i>cīwò</i>                        |
| ‘easy = light’       | <i>néwè</i>  | -             | <i>don</i>   | -                                  |
| ‘cover = close’      | -            | <i>dátugu</i> | <i>daabu</i> | <i>rufē</i>                        |
| ‘top/above = sky’    | <i>kànmú</i> | <i>sán</i>    | <i>beena</i> | <i>samà</i>                        |
| ‘tree = firewood’    | -            | <i>jíri</i>   | <i>tuuri</i> | <i>itàcē</i>                       |
| ‘kill = extinguish’  | <i>kàrí</i>  | <i>fàga</i>   | <i>wii</i>   | <i>kashè</i>                       |
| ‘separate = divorce’ | <i>fátá</i>  | -             | <i>fey</i>   | <i>rabà</i>                        |
| ‘father = FaBr’      | <i>fàabé</i> | <i>fà</i>     | <i>baaba</i> | <i>bàbānī</i> (dialectal, archaic) |

|                    |              |                 |              |  |
|--------------------|--------------|-----------------|--------------|--|
| ‘hard = difficult’ | <i>xote</i>  | <i>gèlɛn</i>    | <i>sendi</i> | -  |
| ‘stranger = guest’ | <i>mùkké</i> | <i>fàrankan</i> | <i>yow</i>   | <i>bàkō</i>                                  |
| ‘mouth = edge’     | <i>ràqè</i>  | <i>dá</i>       | <i>mee</i>   | <i>bàkī</i>                                  |
| ‘hear = smell’     | <i>mùgú</i>  | <i>mén</i>      | <i>maa</i>   | <i>ji</i>                                    |
| ‘foot = time’      | <i>tá</i>    | <i>sèn</i>      | <i>cee</i>   | - ( <i>sau</i> ‘foot’ vs. <i>sàu</i> ‘time’) |

Further west, Western Bade (10/13), Kanuri (9/15), and Mokilko (9/14) are only slightly lower – all spoken near the northern fringes of the Sahel. Further south, the scores are lower again: in the area around Lake Chad, Bole (8/15), Lamang (7/14), Bagirmi (8/14); in the highlands south of the Niger bend, Moore and Tommo So (both 8/14), Bobo (6/13); and, down in Sierra Leone, Kisi (8/13). At the eastern and western fringes of the area, the North Atlantic languages Wolof and Fulani stand out for relatively low scores (5/15), as does (Chadic) Polci (5/14) at the southern edge and the more recently arrived Nigerian Arabic (4/15). This suggests a Sprachbund stretching across the Sahel proper, west to east along the upper and middle Niger and around Lake Chad, where the relative ease of riverine travel facilitated long-distance multilingualism and the spread of lingua francas, percolating more slowly out to the surrounding hinterlands.

Unravelling the history of this Sahelian lexical Sprachbund will require considerably more diachronic work than is feasible in the scope of this article; in principle, comparative etymological studies of Mande and Chadic at least would be required to determine where these colexifications first emerged and how they spread. However, among languages that have entered or left the relevant area within the past millennium, a couple belong to families whose history is relatively well understood: Arabic and Songhay. A detailed examination of these cases reveals the mechanisms at work in practice. Before looking at the etymological details, however, a theoretical excursus is necessary, to examine which mechanisms could in principle make languages come to share colexifications through contact.

#### 4 Diachronic development

How do languages acquire or lose the colexifications characteristic of a particular region? A priori, two mechanisms should be expected. Assume that in language A, the two meanings P and Q are expressed by the same form X, whereas in language B, meaning P is expressed by form Y and meaning Q by form Z (cf. Table 3).

*Table 3. Initial form-meaning mappings in a hypothetical situation.*

|                        |          |          |
|------------------------|----------|----------|
| form in language A     | <i>X</i> | <i>X</i> |
| form in language B     | <i>Y</i> | <i>Z</i> |
| etic meaning (see 2.1) | P        | Q        |

1) *Borrowing* – the adoption of a form along with its associated meaning. If fluent bilingual speakers borrow form Y from language B into language A, they will typically borrow it with its original meaning P. Once X and Y are both available in language A for expressing meaning P, Grice’s Maxim of Manner (1975), aided by the salience of the P-Q meaning distinction for such bilinguals, may tend to favour Y, leaving X with the meaning Q alone and thus eliminating language

A's former colexification of P and Q, as in Table 4. An example in English is the borrowing of *palate* from Latin, causing Old English *gōma* 'gum, palate' to be narrowed to the meaning of its modern reflex *gum* (Millward & Hayes 2012: 205). *A fortiori*, that colexification will also be lost if forms Y and Z are both borrowed into language A with their original meanings and replace form X there. Conversely (though less probably), if fluent bilingual speakers borrow form X from language A into language B, Grice's Maxim of Quantity, aided by the presumed prestige of language A, may favour the loss of both Y and Z, causing language B to colexify the previously distinct meanings P and Q as in Table 5. An example in English is the borrowing of *uncle* from French, replacing Old English *fædera* 'paternal uncle' and *ēam* 'maternal uncle' (Pfeffer 1987) and accompanied by the complete loss of this previously very salient distinction. In either case, lexical borrowing leads to lexical restructuring.

Table 4. Results of borrowing from B to A

|                    |          |   |
|--------------------|----------|---|
| form in language A | <b>Y</b> | X |
| form in language B | Y        | Z |
| etic meaning       | P        | Q |

Table 5. Results of borrowing from A to B

|                    |          |          |
|--------------------|----------|----------|
| form in language A | X        | X        |
| form in language B | <b>X</b> | <b>X</b> |
| etic meaning       | P        | Q        |

2) *Calquing* - the analogical application of a structural principle without accompanying form. Bilingual speakers fluent in language A will often be tempted to express both P and Q with a single form Y when speaking language B, using a form belonging to that language to express a meaning that would previously have been expressed with a different form or not at all. As a result, language B gains the colexification of meanings P and Q, previously not colexified there, as in Table 6. A late 20<sup>th</sup> century example in French is the extension of *souris* 'mouse (animal)' to express 'mouse (computer peripheral)', on the model of English *mouse*. The converse situation, while less economical, is also possible. If the distinction between P and Q is sufficiently salient for speakers of language A, then bilingual speakers may feel the need to make that distinction in language B. This is most easily achieved by borrowing, but when language ideologies strongly disfavour borrowing, it can also be achieved using internal morphological resources by creating a neologism (form N) to express Q, and relegating form X to the meaning of P alone, as in Table 7. As an example of this probably less common process, one may cite the split of Japanese *kare*, which in the early 19<sup>th</sup> century still meant 's/he, that person', into *kare* 'he' and the newly coined form *kanojo* 'she' as a result of translations from Western languages (Miura 1979; Yanabu 2011). In either case, lexical calquing leads to lexical restructuring.

Table 6. Results of calquing from A to B

|                    |   |          |
|--------------------|---|----------|
| form in language A | X | X        |
| form in language B | Y | <b>Y</b> |
| etic meaning       | P | Q        |

Table 7. Results of calquing from B to A

|                    |   |   |
|--------------------|---|---|
| form in language A | X | N |
| form in language B | Y | Z |
| etic meaning       | P | Q |

While the dichotomy of borrowing vs. calquing laid out here is close to what Matras and Sakel (2007) label *matter* vs. *pattern* borrowing, the parallel is not exact; matter borrowing can affect lexical structure only to the extent that it implies pattern borrowing as well. When the speakers are significantly less fluent in the source language than in the recipient language, they may well borrow a form but use it only to express the meaning of a lexical item already present in the recipient language, as in Lefebvre’s (2014) “relabelling”.

By looking at Nigerian Arabic and Korandje in the context of the fairly well-documented families to which they belong, it is possible to evaluate the relative importance of these two mechanisms.

#### 4.1 In Nigerian Arabic

Arabic, originating in southwest Asia, expanded into northern Africa starting in the 7<sup>th</sup> century CE. From Egypt, speakers of Arabic moved up the Nile into modern-day Sudan, and thence westward into the Lake Chad area starting around the 14<sup>th</sup> century. Two of the colexifications shared by Nigerian Arabic on this list seem likely to predate its arrival in the area:

- *īd* ‘hand = arm’ < Classical and Sudanese Arabic *yad* ‘hand (palm); hand+arm (from fingertips to shoulder)’ (Qāsim 1972), with no single word for the arm excluding the hand. Most Arabic dialects allow reflexes of *dirāʿ* ‘forearm’ to be used to refer to the arm as a whole, often (like Algerian Arabic) even while retaining the colexification of ‘hand = hand+arm’. However, full-blown ‘hand = arm’ colexification is fairly frequent, particularly in the Levant (Behnstedt & Woidich 2010: 143). The same colexification *īd* ‘hand = arm’ is found in Chadian Arabic (Jullien de Pommerol 1999).
- *gāsi* ‘hard (surface) = difficult’ < Classical *qāsi(n)* ‘hard (surface)’, occasionally also ‘difficult’. Sudanese Arabic retains the primary meaning ‘tough, hard’ but uses the plural *gawāsī* in the sense of ‘difficulties’, while *gāsi* is ‘difficult, expensive’ in Chadian Arabic.

Since its arrival in the Lake Chad region, however, Nigerian Arabic has acquired three newer colexifications characteristic of the area but otherwise less frequently documented in Arabic.

- *dēf* ‘stranger = guest’ < Classical *ḡayf* ‘guest’, with reflexes preserved in the sense of ‘guest’ throughout almost all modern Arabic varieties (Behnstedt & Woidich 2010: 59), and contrasting with a variety of near-synonyms for ‘stranger’, such as *yarīb*. This colexification is not reported for Sudanese Arabic, but is found in Chadian Arabic *dēf*.
- *qašim* ‘mouth = edge’ < Classical *xašm* ‘cartilage of the nose’, via the sense it had already acquired in Upper Egyptian and Sudanese Arabic long before reaching Nigeria, ‘mouth’ (Behnstedt & Woidich 2010: 125). This seems to be a particularly recent development; both Sudanese and Chadian *xašum* are only reported as having the senses ‘mouth, entrance’, and both retain distinct words for ‘edge’ (*ṭaraf* / *taraf* respectively.)



- *ríjile* ‘feet = times (instances)’ < Classical *rijl* ‘leg’; not documented by Kaye (possibly a later innovation?) but discussed in Owens (1993:224-226). The two senses are colexified in the plural, but are distinguished in the singular by the Arabic feminine / singulative / plural suffix *-e*: *rijil* ‘foot’ vs. *ríjile* ‘time’. Since the singular remains distinct, this partial colexification is not counted on the maps.

The strongest immediate influence on Nigerian Arabic is from the regionally dominant language Kanuri, which has *músko* ‘hand = arm’, *kusóto* ‘stranger = guest’, *cî* ‘mouth = bank, shore, side’, *shî* ‘foot = time’. While the first one may well be a coincidence, the latter three appear good candidates for having been calqued into Nigerian Arabic. However, the region is sufficiently multilingual that multiple causation involving other languages of the area cannot be excluded. What is clear is that none of the colexifications observed on this list involved material borrowing; apart from accidental convergence, the only relevant mechanism is calquing. The few changes observed have thus had the effect of requiring fewer distinct lexemes and more contextual interpretation. Such changes constitute simplification by the definition of Trudgill (2009), in that it is presumably easier for an arbitrary L2 learner to guess an unfamiliar contextual interpretation of a known lexeme than to guess the meaning of an unknown lexeme (and, in this specific context, undoubtedly easier for a Kanuri speaker). Following van Coetsem’s (1988) model, we may expect that they were imposed by Kanuri-dominant speakers; but since Kanuri is the lingua franca of the region, such speakers – until the onset of large-scale urbanization – would have mainly been outsiders integrating into Arab communities, poorly positioned to persuade others to adopt their way of speaking. The absence of borrowing in this context contrasts strikingly with the situation in Songhay seen below.

#### 4.2 In Korandje and the rest of Songhay

Gao, by construction, has all 15 colexifications; Timbuktu has 14, Zarma 13, Hombori 12-13, Tasawaq 8-11. Yet Tadaksahak has only 5-8, and Korandje only 2. In order to account for this striking difference, it is necessary to reconstruct the history of these colexifications within the family.

Phylogenetically (Souag 2012), Songhay divides naturally into a not necessarily coherent, relatively conservative “eastern” dialect continuum, mostly along the Niger River but with outliers to its south, and a somewhat restructured Northwestern subgroup with SVO basic order; the latter in turn divides into a lexically conservative Western subgroup along the Niger River, and a heavily Berber-influenced Northern subgroup scattered through the Azawagh region of northern Niger, with one outlier in western Algeria. The structure is as follows:

- Eastern: Kikara, Hombori, Gao (Mali), Kaado (Gorouol, Niger), Zarma (Niger), Dendi (Djougou, Benin)
- Northwestern:
  - Western: Djenne, Timbuktu
  - Northern: Korandje (Tabelbala, Algeria), Tadaksahak (Mali/Niger), Tagdal (Niger), Tasawaq (In-Gall, Niger), Emghedesie (Agades, Niger – extinct)

Plausible criteria for reconstructing a colexification are that it should be shared across most varieties and that, where the distinction is made, it is made using a form that cannot be reconstructed for

proto-Songhay, either a clear loanword or a different reflex of the same etymon. By these criteria, at least twelve of the fifteen colexifications appear reconstructible for proto-Songhay.

1. *\*ndá* ‘with (comitative) = with (instrumental)’; the same polysemy is shared by Zarma and Kikara *dá*, Hombori *ndù*, Tadaksahak *əndá*, Tagdal *nda*, and, to an extent, Djenne, Timbuktu and Gao *nda*, and is securely reconstructible for proto-Songhay. Korandje, however, has split this into *ndza* ‘with (comitative)’ vs. AGR+*indza* ‘with (instrumental)’ through metanalysis of a bipartite comitative structure calqued from Berber (Souag 2015b).
2. *\*kàmbè* ‘hand = arm’, like Djougou *kàmmè*, Zarma *kàmbè*, Gorouol *kàbè*, Hombori *kàmbè*, Gao *kabe*, Timbuktu *kamba*. (In Djenne and Kikara, no word for ‘arm’ is documented, only ‘hand’, respectively *kamba*, *kàmbè*.) Even though Tuareg distinguishes *əfús* ‘hand’ from *ayil* ‘arm’, this is stable in the Azawagh languages; Agades *kánba* ‘hand; upper part of the arm’, In-Gall *kâmbà* ‘arm (including the hand)’, Tagdal *kāmba* ‘arm, hand’, Tadaksahak *kambá* ‘hand, arm’. Korandje is unique within Songhay in distinguishing *kambi* ‘hand, sleeve, handle’ from *ddrəŋ* ‘arm, cubit’; the latter derives transparently from Maghrebi Arabic *draŋ* < Classical Arabic *dirāŋ* ‘forearm’.
3. *\*dóórú* ‘hurt’ (but *dóor-í* ‘pain = illness’) vs. *\*záŋáy* = ‘be (seriously) ill’; these were probably distinct in proto-Songhay, but their reconstruction is not trivial. *\*dóórú* ‘hurt’ is fairly secure, being synchronically attested as such in Djenne *dɔɔr*, Timbuktu *door*, Gao *dor*, Gorouol *dóré*, Zarma *dóórú*, Djougou *dóórù*, Tadaksahak *dâr*, Korandje *ḍa*. Only in Gao has a reflex gained the sense ‘be ill’, which accordingly should not be reconstructed for this verb. The corresponding noun *\*dóor-í*, however, does seem to have colexified ‘illness’ and ‘pain’: Gao *doori* ‘harm, pain; cut, open sore, wound’, Zarma *dóóri* ‘pain, illness, suffering’, Djougou *dóóri* ‘pain’, *dâr* ‘sickness, epidemic, suffering’, In-Gall *dóóri* ‘be ill, illness’, Agades <*dóri*> ‘ill, illness’. For the verb ‘be ill’, we must rather look to *\*záŋáy*: Gao *zaney* ‘be ailing, be hurt’, Gorouol *záŋéy* ‘be ill’, Zarma *záŋáy* ‘be ill bedridden’, Kikara *záy* ‘(body part) hurt, (sb) be sick or injured’, Djougou *záŋŋáy* ‘illness’. In Northwestern Songhay and Gao, we instead find reflexes of *\*wirki* > *wirči* ‘be (very) sick, be a medical patient’, perhaps an early loan from dialectal Arabic *hlk* ‘be ill’. Zarma *jántè* ‘be ill, have a fever’ seems isolated within Songhay, while Tadaksahak *hé akkúna*, lit. “something found him” = “he is ill”, is a partial calque from Tamasheq *tagraw-tu torhanna* “an illness found him”. It thus appears that, whereas the nominal colexification ‘pain = illness’ is old in Songhay, the corresponding verbal colexification ‘hurt = be ill’ is a later secondary development in Korandje, Gao and, from a different root, in Kikara. The former internally motivates the latter, independently of – or in addition to – contact.
4. *\*dògón* ‘easy = light (weight)’, as in Gao *don*, Gorouol *dóon*, Hombori *dò:n*, Zarma *dògón*; in Kikara, a trace remains in the frozen causative *dó:n-ân* ‘make light(-weight); make easy’, although Fulani borrowings have created a distinction between *hóyè* ‘easy’ and *hóyfidè* ‘be or become lightweight’. In Djenne *baan* ‘be light, easy; be soft’, Timbuktu *baan* ‘be soft; be easy; be light’, the original meaning ‘soft’ (cp. Gao *ban* ‘be soft, tender’) has been extended to the other two senses, maintaining the older colexification intact but replacing the form. For Agades, both meanings are expressed in available data as negatives: <*ni-tshíng*> ‘light (not heavy)’ vs. <*ne-géb*> ‘easy, cheap (not hard, dear)’. For ‘light’, In-Gall borrows *fùsús* ‘be light’ from Berber; for ‘easy’, In-Gall uses a negative expression based on Hausa, *ní-*

*mày wúyá* ‘not have difficulty’. Korandje contrasts the Berber borrowing *fusəs* ‘light’ with the Arabic borrowing *yəhwən* ‘easy’.

5. \**dáabù* ‘cover = close’, like Djenne, Timbuktu, Gao *daabu*, Gorouol *dáabù*, Hombori *dá:bù*, Zarma and Djougou *dáabù*. In the Azawagh, this is maintained in Tadaksahak *dáb* (Heath 2007), Tagdal *dəb/dab*, Agades *dab* (assuming that <*dabá nga*> “I put out, extinguish” was rather “I cover it”.) In In-Gall, *dáb* ‘close’ alongside *dábànhà* ‘top, stopper (*bouchon*)’ makes it clear that this polysemy was originally found. Korandje retains *dəb* ‘cover’, but – uniquely within Songhay – for ‘close’ uses *həw* < ‘tie’, calquing a widespread Berber polysemy exemplified by Tashelhiyt *qqn* ‘close, tie’.
6. \**béené* ‘top/above = sky’, like Zarma *béené*, Gorouol *béené*, Hombori *bé:né*, Gao *beena*, Timbuktu *beene*, Djenne *béenə*. In the Azawagh, this colexification is stable in the sedentary varieties; thus Agades *béne* ‘on, over’, *bene* ‘upper story, tower’, *bíní* ‘heaven’ (all attempts to transcribe the same word), In-Gall *béené* ‘ciel ; sur, au-dessus’. In the nomadic varieties, the colexification is avoided by borrowing a Tuareg word for ‘sky’: Tadaksahak *beená* ‘top’ vs. *izínnan* ‘sky’, Tagdal *iggunən* ‘ciel’. Korandje preserves the memory of this colexification in a popular line of macaronic doggerel *əl-ṣayš ṭazu, əl-ma iri, əl-ərḍ gandza s-sma bini* (“couscous is *ṭazu*, water is *iri*, the earth is *gandza*, the sky is *bini*”). In normal usage, however, *bini* is indeed ‘above, roof’, but ‘sky’ is Berber *igərwan*.
7. \**tùgúdì* ‘tree = firewood’, like Djenne *tuguri*, Timbuktu and Gao *tuuri*, Hombori *tù:rî*, Zarma and Djougou *túúrì*. (Gorouol *túuri* ‘arbre, médicament’ is not precisely enough glossed.) Kikara *tũ:rî* ‘wood, firewood’ is distinct from *həʔ* ‘tree’, with no known cognates. Tadaksahak *tugúdu* ‘tree, wood’, Tagdal *tuguzu* ‘firewood; tree’, Agades *tugújio* ‘tree; firewood’, and In-Gall *tùgúzi* ‘tree, wood, post’ all maintain this polysemy. Korandje, however, splits it into *skudzi* ‘(fire)wood’ and the Arabic loanword *aṣṣaḍṛaṭs* ‘tree, woody shrub’.
8. \**wí* ‘kill = extinguish’, like Djougou *wí*, Zarma *wí*, Gorouol *wí*, Hombori *wí*: ~ *wúy*, Gao *wii* ‘kill; extinguish (e.g. light)’, Timbuktu *wii*, Djenne *wii*. (For Kikara, no word for ‘extinguish’ is recorded; only *wúy* ‘kill’.) The Azawagh languages for which data is available preserve this colexification: In-Gall *wí*, Tadaksahak *wí*. So does Korandje *wi*.
9. \**fáy* ‘separate = divorce’, like Djenne *fay*, Timbuktu and Gao *fey*, Gorouol and Hombori *féy*, Zarma *fáy*. (Kikara *fáy* ‘separate’ is distinct from *sé:rè* ‘divorce’.) In-Gall seems to maintain the colexification of *fáyfây* ‘be divorced; separate’. However, Tadaksahak *fáy* ‘to be separated; to separate’ is distinct from the Berber loanword *léf* ‘divorce’. Korandje has likewise abolished the distinction, while losing \**fáy* entirely: ‘separate’ is the Arabic loan *fagəm*, while *fya* ‘release, divorce’ copies the polysemy of Arabic *ṭlq*.
10. \**bàabà* ‘father = father’s brother’, like Zarma *bààbà*, Gorouol *bàabè*, Hombori *bà:bè*, Gao and Timbuktu *baaba*. The same colexification is clearly evidenced, at least historically, in Djougou *bààbà* ‘father’, *à-bààbé* ‘my father’, *bààbé béérì / tyéénè* (‘big’/‘little’) ‘elder/younger paternal uncle’. (Available materials suggest that Kikara and Djenne contrast *bèbè / baaba* ‘father’ with *bà-kàynè/bè:rì / baa-kayna/bèer* ‘father’s younger/elder brother’; however, the latter terms are well-attested in languages where reflexes of \**bàabà* unambiguously do show this colexification, and this may simply be a documentation gap.) In the Azawagh, the colexification (shared with Tuareg) has left clear traces; we have Tadaksahak *baabá* ‘father; brother of the father; ancestor’ alongside a dedicated term for

‘paternal uncle’, *mošaddí* (Chaventré 1983; Souag 2015c); In-Gall *báàbà* ‘father’, *bààbà káyà* ‘uncle’. In Korandje, no such colexification is attested; *abba* ‘father’ < Berber *abba* contrasts sharply with *šammi* ‘paternal uncle’ < Berber/Arabic *šammī*.

11. *\*séndí* ‘hard (surface)’ vs. *\*gáábù* ‘difficult, refractory’ (with its more widely attested corresponding noun *\*gáab-í* ‘strength, power’). The colexification ‘hard = difficult’ is universal in Songhay varieties spoken along the river, found in Gao *sendi*, Kikara *séndí*, Zarma *sándí* and *gáábù* (probably also in Djénne and Timbuktu *sendu* and Gorouol *séndí*, though the glosses are a little ambiguous.) Off the river, however, the distinction is often maintained: Hombori contrasts *séndí* ‘be difficult, expensive’ with *jítí* ‘hard’, while Djougou contrasts *sènní* ‘hard (*dur*), solid, petrified’ with *gáábù* ‘difficult (be)’. In Northern Songhay, this colexification is likewise unattested. Agades contrasts <*shéndi*> ‘hard, dry’ with <*géb*> ‘difficult, dear’, and In-Gall contrasts *sándi* ‘be hard (*dur*)’ with the partial Hausa loan *mày wúyá* lit. ‘have difficulty’ (alongside *gáb* ‘be expensive’.) Korandje has lost *\*séndí*, but contrasts *kəʃ* ‘hard’ (< *\*kású* ‘rough’) with *gəb* ‘difficult’.
12. *\*yèw* ‘outsider = guest’, like Djougou *yò*, Zarma *cè-yàw*, Gorouol *yèw*, Hombori *yòw-ò*, Kikara *yòw*. This root does not appear to have survived anywhere in Northern Songhay. The Azawagh languages have borrowed Tuareg *amāgar* ‘guest, outsider’, probably keeping the colexification; Agades *amāgar* ‘guest’, In-Gall *àmágàr* ‘étranger’, Tagdal *amageru* ‘étranger’. Korandje, however, has borrowed distinct words for both concepts: *ḍḍayf* ‘guest’ < Arabic *ḍayf*, *abərrən* ‘stranger’ < Berber *a-bərran* < Maghrebi Arabic *bərrani*.
13. *\*míyè* ‘mouth = edge’, like Djougou *mée*, Zarma *mê*, Kaado *mè*, Gao, Timbuktu *mee*, Djénne *mεε* ‘mouth’. In the Azawagh, both senses are at least sometimes retained, as in In-Gall *mè* and possibly Tadaksahak *míya* ‘mouth, peak, muzzle; end, side’. In Korandje, *mi* ‘mouth; entrance; mouth (of pot)’ has lost this sense; ‘edge’ is rather expressed with *alḥašiyət* < Arabic *ḥāšiyah*.
14. *\*máwá* ‘hear = smell (v.)’, like Djougou, Zarma, Gorouol *má*, Hombori *má:*, Gao *maa*, Timbuktu *mom* / *mow*, Djénne *mɔm* / *mɔ*. This colexification is maintained in In-Gall *máw*, but lost in Tadaksahak *mó* ‘to hear (noise); to have news about sth’ vs. *kár mammaní* ‘to smell (odour)’ (*kár* ‘to hit’), calqued from Tuareg *əwət aḍu* ‘hit wind’ = ‘smell’). Korandje has *məw* ‘hear’ versus *išəmm* ‘smell’ < Arabic.
15. *\*kèe* ‘foot’, *\*ké* ‘time (instance)’ poses greater difficulty; were the two forms distinct in proto-Songhay or not? The tonal distinction – the inverse of Hausa *sau* ‘foot’ vs. *sàu* ‘time’ – is made in widely separated varieties, cf. Djougou (Heath 2001) *čè*: ‘foot’ vs. *čé*: ‘time’, Zarma *cè* vs. *cé*, Kikara *kè*: vs. *ké*. However, two other varieties distinguish them only through clearly or probably secondary developments: Hombori *čè*: vs. *čè*, Djénne *čee* vs. *čee*. The distinction is absent further north, even in one variety that retains tone: Gorouol *cè*, Gao and Timbuktu *čee* have both senses. It seems most likely that a single word with two senses (as found in Bambara and Kanuri) developed different tones in proto-Songhay due to contextual factors. Reconstructing the two as coincidentally similar but originally distinct etymons would force one to suppose that the colexification of these meanings in Kanuri, Moore, Soninke, Bambara, Bole, Bagirmi, etc, was ultimately calqued on a Gao-like Songhay variety where the two had merged, which seems problematic given the lack of evidence for Songhay influence as far east as Kanuri or Bagirmi. In Agades, <*ké*> ‘leg’

may or may not be homophonous with <kae> ‘time’, while Tasawaq unambiguously differentiates *sè* ‘leg, foot’ from *sô* ‘time (*fois*)’ (whose unexpected vowel is probably copied from its common companion *fó* ‘one’). Korandje, however, has replaced ‘time’ with *tsara* < Berber *tawala*, while retaining *tsi* ‘foot, leg’.

Looking at this data, we see that the influence of neighbouring languages has operated in three distinct ways to erase former colexifications:

1. borrowing of words expressing one or both of the senses (Korandje ‘hand’ vs. ‘arm’, ‘easy’ vs. ‘light’, ‘top’ vs. ‘sky’, ‘tree’ vs. ‘firewood’, ‘separate’ vs. ‘divorce’, ‘father’ vs. ‘father’s brother’, ‘outsider’ vs. ‘guest’, ‘mouth’ vs. ‘edge’, ‘hear’ vs. ‘smell’, possibly ‘foot’ vs. ‘time’; Tadaksahak ‘top’ vs. ‘sky’, ‘separate’ vs. ‘divorce’, ‘hear’ vs. ‘smell’, possibly ‘hurt’ vs. ‘ill’)
2. calquing of Arabic/Berber polysemy to express one or both of the senses using a different inherited etymon (Korandje ‘cover’ vs. ‘close’, ‘separate’ vs. ‘divorce’)
3. metanalysis of an existing etymon in one sense due to calquing of an Arabic/Berber construction (Korandje ‘instrumental’ vs. ‘comitative’)

Out of borrowing, calquing, and metanalysis, borrowing seems to be by far the most important factor in the adaptation of Korandje to North African lexical typology and of Tadaksahak to Tuareg lexical typology over the past millennium or less. This has restructured parts of the lexicon in a manner requiring more distinct lexemes and less contextual interpretation. Such a change constitutes complexification by the definition of Trudgill (2009), in that it is presumably easier for an arbitrary L2 learner to guess an unfamiliar contextual interpretation of a known lexeme than to guess the meaning of an unknown lexeme (though not for a speaker of the language it was borrowed from!), and in that the latter strategy increases redundancy by providing additional disambiguation on top of what context had already provided. This change is thus consistent with the generalization that complexification predominates in small isolated communities (e.g. Wray & Grace 2007; Trudgill 2009); Korandje has only a few thousand speakers, concentrated in an oasis 100 km away from any other settlements. Yet, contrary to common expectations, this complexification has been driven largely by contact, suggesting – despite clear evidence for immigrants learning Korandje as L2 – that (as is the case today) Korandje speakers were more frequently learning other languages (essential for any kind of travel beyond the oasis) than vice versa.

Despite the important role of borrowing in eliminating colexifications, three colexifications that do not seem to be securely reconstructible for proto-Songhay are nevertheless well-attested: ‘hurt = ill’ (Korandje, Gao, Kikara), ‘hard = difficult’ (Gao, Kikara, Zarma), ‘foot = time’ (Gorouol, Gao, Timbuktu). All three are shared with Bambara to the west, which has *bàna* ‘hurt = ill’, *gèlèn* ‘hard = difficult’, *sèn* ‘foot = time’; one is found in Hausa to the east as well, which has *cīwò* ‘hurt = ill’. The phonetic forms involved, however, are unambiguously inherited from proto-Songhay in all these cases. In each case individually, internal development could provide a sufficient account for the emergence of colexification in specific languages, aided by the corresponding noun for ‘hurt = ill’ and by phonetic similarity for ‘foot = time’ (if the latter two are assumed to have been originally distinct). Given the bigger picture, however, contact appears likely to have facilitated the process, in this case through calquing rather than through borrowing.

Korandje and Nigerian Arabic thus show about the same amount of calquing on the items in this list, while contrasting sharply in the extent of borrowing. The difference in their level of convergence to regional lexical-typological norms here results almost entirely from the difference in borrowing.

## 5 Conclusions

This paper demonstrates for the first time that the languages of the western Sahel/Savanna group with Songhay as against North Africa in lexical-typological terms where the two differ, and thus constitute a lexical Sprachbund, facilitated by the relative ease of east-west travel (perhaps especially along the Niger and Yobe Rivers.) Languages leave or join this lexical Sprachbund primarily through two mechanisms familiar from contact linguistics: calquing, and borrowing.

In principle, both borrowing and calquing may have similar effects on a language's lexical typology (see the beginning of Section 4), but their roles in this region have been quite different. Within the data set examined here (including the less contact-intense cases in Section 3.1), borrowing usually causes the loss of colexifications, while calquing consistently causes them to be gained, emerging as the key mechanism for gaining the colexifications shared across the western Sahel/Savanna. It remains to be determined whether this can be generalized worldwide – and, if not, which factors have caused it to hold true around the Sahara.

The data examined here also demonstrates that these two principal mechanisms of convergence can operate independently at very different speeds. The relatively limited borrowing found in Nigerian Arabic is simply not comparable to the massive borrowing that, over the same period of time, has left Korandje with only a few hundred (highly frequent) words of Songhay origin. The principal explanation for this difference is no doubt sociolinguistic: Nigerian Arabic has far more speakers (hundreds of thousands, vs. Korandje's few thousand), and seems to enjoy somewhat more prestige. Yet, for the colexifications examined, calquing has proceeded equally slowly in both languages, affecting only 2-3 out of 15 items in the course of about 700 years. Given the major differences between the two cases, this suggests that the rate of calquing in basic vocabulary may be less variable cross-linguistically than the rate of borrowing, as might be expected from borrowing's greater accessibility to conscious awareness and manipulation. The method of comparing language islands within an area to their relatives outside it should make it possible to investigate this question on a broader scale.

## References

- Behnstedt, Peter & Manfred Woidich. 2010. *Wortatlas der Arabischen Dialekte. Band I: Mensch, Natur, Fauna und Flora*. Leiden: Brill.
- Brown, Cecil H. 2013. Hand and Arm. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Max Planck Digital Library. <http://wals.info/chapter/129> (31 January, 2019).
- Brown, Francis, Samuel R. Driver & Charles A. Briggs. 1906. *A Hebrew and English Lexicon of the Old Testament with an appendix containing the Biblical Aramaic*. Oxford: Clarendon Press.
- Cancel, Lt. 1908. Etude sur le dialecte de Tabelbala. *Revue Africaine* 270–271. 302–347.
- Caron, Bernard & Petr Zima (eds.). 2006. *Sprachbund in the West African Sahel*. Louvain: Peeters.
- Chaker, Salem & Dominique Caubet (eds.). 1996. *La négation en berbère et en arabe maghrébin*. Paris: L'Harmattan.

- Chaventré, André. 1983. *Evolution anthropo-biologique d'une population touarègue: Les Kel Kummer et leurs apparentés* (Travaux et Documents Cahier no. 103). Paris: Presses universitaires de France.
- Clements, G.N. & Annie Rialland. 2008. Africa as a phonological area. In Bernd Heine & Derek Nurse (ed.), *A Linguistic Geography of Africa* (Cambridge Approaches to Language Contact), 36–85. Cambridge University Press.
- Dryer, Matthew S. 1989. Large Linguistic Areas and Language Sampling. *Studies in Language* 13(2). 257–292. doi:10.1075/sl.13.2.03dry.
- Fentress, Elizabeth. 2009. Slavers on Chariots: The Garamantes between Siwa and the Niger Bend. In Amelia Dowler & Elizabeth R. Galvin (eds.), *Money, Trade and Trade Routes in Pre-Islamic North Africa*. London: British Museum.
- François, Alexandre. 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In Martine Vanhove (ed.), *Studies in Language Companion Series*, vol. 106, 163–215. Amsterdam: John Benjamins.
- Greenberg, Joseph H. 1963. *The Languages of Africa*. Bloomington: Indiana University.
- Grice, George. 1975. Logic and conversation. In P. Cole & J. Morgan (eds.), *Syntax and Semantics. 3: Speech Acts*, 41–58. New York: Academic Press.
- Güldemann, Tom. 2018. Language contact and areal linguistics in Africa. In Tom Güldemann (ed.), *The Languages and Linguistics of Africa*, 445–545. Berlin: De Gruyter Mouton. doi:10.1515/9783110421668-003.
- Haspelmath, Martin & Uri Tadmor (eds.). 2009. *Loanwords in the world's languages: a comparative handbook*. New York, N.Y: Mouton de Gruyter.
- Hayward, Richard J. 1991. A propos patterns of lexicalization in the Ethiopian Language Area. In Daniela Mendel & Ulrike Claudi (eds.), *Ägypten im afroorientalischen Kontext. Special issue of Afrikanistische Arbeitspapiere*, 139–156. Cologne: Institute of African Studies.
- Hayward, Richard J. 2000. Is there a metric for convergence? In Colin Renfrew, April McMahon & Larry Trask (eds.), *Time Depth in Historical Linguistics*, vol. 1, 621–640. Cambridge: McDonald Institute for Archeological Research.
- Heath, Jeffrey. 1998. *Dictionnaire Songhay-Anglais-Français* (Langues d'Afrique 4–6). Paris: L'Harmattan.
- Heath, Jeffrey. 2001. Dendi lexicon. Unpublished ms. [http://www-personal.umich.edu/~jheath/Dendi\\_lex.pdf](http://www-personal.umich.edu/~jheath/Dendi_lex.pdf).
- Heath, Jeffrey. 2007. Tadaksahak lexicon. Unpublished ms.
- Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.
- Heine, Bernd & Zelealem Leyew. 2007. Is Africa a linguistic area? In Bernd Heine & Derek Nurse (eds.), *A Linguistic Geography of Africa*, 15–35. Cambridge: Cambridge University Press.
- Jullien de Pommerol, Patrice. 1999. *Dictionnaire arabe tchadien-français : suivi d'un index français-arabe et d'un index des racines arabes*. Paris: Karthala.
- Kaye, Alan S. 1982. *A dictionary of Nigerian Arabic* (Bibliotheca Afroasiatica 1). Malibu, CA: Undena.
- Kaye, Alan S. 1987. *Nigerian Arabic-English dictionary* (Bibliotheca Afroasiatica 2). Malibu, CA: Undena.
- Koptjevskaja-Tamm, Maria & Henrik Liljegren. 2017. Lexical semantics and areal linguistics. In Raymond Hickey (ed.), *The Cambridge Handbook of Areal Linguistics*, 204–236. Cambridge: Cambridge University Press.
- Kossmann, Maarten. 1997. *Grammaire du parler berbère de Figuig (Maroc oriental)*. Paris: Peeters.
- Lefebvre, Claire. 2014. *Relabeling in language genesis*. New York: Oxford University Press.
- List, Johann-Mattis, Simon Greenhill, Cormac Anderson, Thomas Mayer, Tiago Tresoldi & Robert Forkel (eds.). 2018. *CLICS<sup>2</sup>*. Jena: Max Planck Institute for the Science of Human History. <https://clics.clld.org/> (12 April, 2019).

- Matras, Yaron & Jeanette Sakel. 2007. Investigating the mechanisms of pattern replication in language contact. *Studies in Language* 31(4). 829–865.
- Michailovsky, Boyd. 2002. *Limbu-English dictionary of the Mewa Khola dialect with English-Limbu index*. Kathmandu: Mandala Book Point.  
<https://lacito.vjf.cnrs.fr/pangloss/dictionaries/limbu/index.html>.
- Millward, C. M. & Mary Hayes. 2012. *A Biography of the English Language*. 3rd ed. Boston: Wadsworth.
- Miura, Akira. 1979. The Influence of English on Japanese Grammar. *The Journal of the Association of Teachers of Japanese*. American Association of Teachers of Japanese 14(1). 3–30.  
doi:10.2307/489539.
- Morgan, Lewis Henry. 1871. *Systems of consanguinity and affinity of the human family*. Washington, D.C.: Smithsonian Institution. (31 January, 2019).
- Murdock, George Peter. 1981. *Atlas of world cultures*. Pittsburgh: University of Pittsburgh Press.
- Muysken, Pieter. 2008. *From Linguistic Areas to Areal Linguistics* (Studies in Language Companion Series 90). Amsterdam: John Benjamins.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago ; London: University of Chicago Press.
- Olson, David M., Eric Dinerstein, Eric D. Wikramanayake, Neil D. Burgess, George V. N. Powell, Emma C. Underwood, Jennifer A. D'amico, et al. 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience* 51(11). 933–938. doi:10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2.
- Owens, Jonathan. 1993. *A grammar of Nigerian Arabic*. Wiesbaden: Harrassowitz.
- Owens, Jonathan. 2015. Idioms, Polysemy, and Context: A Model Based on Nigerian Arabic. *Anthropological Linguistics* 57(1). 46–98. doi:10.1353/anl.2015.0004.
- Owens, Jonathan & Jidda Hassan. 2004. Remarks on Ideophones in Nigerian Arabic. In Martine Haak, Rudolf Erik de Jong & Kees Versteegh (eds.), *Approaches to Arabic Dialects: A Collection of Articles Presented to Manfred Woidich on the Occasion of His Sixtieth Birthday*, 207–220. Leiden: Brill.
- Owens, Jonathan & Jidda Hassan. 2008. West Sudanic Arabic. (Ed.) Kees Versteegh, Mushira Eid, Alaa Elgibali, Manfred Woidich & Andrzej Zaborski. *Encyclopedia of Arabic Language and Linguistics*. Leiden: Brill.
- Pfeffer, Georg. 1987. The Vocabulary of Anglo-Saxon Kinship. *Homme* 27(103). 113–128.  
<https://doi.org/10.3406/hom.1987.368859>.
- Qāsim, ʿAwn al-šarīf. 1972. *Qāmūs al-Lahjah al-fāmmiyyah fī al-Sūdān*. Beirut: Dār al-Fikr.
- R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Schapper, Antoinette, Lila San Roque & Rachel Hendery. 2016. Tree, firewood and fire in the languages of Sahul. In Maria Koptjevskaja-Tamm & Päivi Juvonen (eds.), *Lexico-typological approaches to semantic shifts and motivation patterns in the lexicon*. Berlin: Mouton de Gruyter. <https://repository.ubn.ru.nl/handle/2066/159843> (15 May, 2019).
- Smith Stark, Thomas C. 1994. Mesoamerican calques. I. In Carolyn J. MacKay & Verónica Vázquez (eds.), *Investigaciones lingüísticas en Mesoamérica*, 15–51. México: Universidad Nacional Autónoma de México.
- Souag, Lameen. 2010. *Grammatical Contact in the Sahara: Arabic, Berber, and Songhay in Tabelbala and Siwa*. School of Oriental and African Studies, University of London.
- Souag, Lameen. 2012. The subclassification of Songhay and its historical implications. *Journal of African Languages and Linguistics* 33(2). 181–213.
- Souag, Lameen. 2013. Sub-Saharan lexical influence in North African Arabic and Berber. In Mena Lafkioui (ed.), *African Arabic: Approaches to Dialectology*, 211–236. Berlin: Mouton de Gruyter.



- Souag, Lameen. 2015a. Explaining Korandjé: Language contact, plantations, and the trans-Saharan trade. *Journal of Pidgin and Creole Studies* 30(2). 189–224.
- Souag, Lameen. 2015b. How to make a comitative preposition agree it-with its external argument: Songhay and the typology of conjunction and agreement. In Paul Widmer, Jürg Fleischer & Elisabeth Rieken (eds.), *Agreement from a diachronic perspective*, 75–100. Berlin: De Gruyter.
- Souag, Lameen. 2015c. Non-Tuareg Berber and the Genesis of Nomadic Northern Songhay. *Journal of African Languages and Linguistics* 36(1). 121–143.
- Souag, Lameen. 2017. La diffusion en berbère : Reconcilier les modèles. In *Diffusion, implantation, convergence* (Mémoires de la Société Linguistique de Paris (Nouvelle Série) 24), 83–107. Louvain: Peeters.
- Stolz, Thomas, Cornelia Stroh & Aina Urdze. 2011. Comitatives and Instrumentals. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Max Planck Digital Library. <http://wals.info/chapter/52> (20 June, 2012).
- Swadesh, Morris. 1952. Lexico-Statistical Dating of Prehistoric Ethnic Contacts. *Proceedings of The American Philosophical Society* 96(4). 452–463.
- Tadmor, Uri. 2009. Loanwords in the world's languages: Findings and results. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the World's Languages: A Comparative Handbook*, 55–75. Berlin: De Gruyter.
- Tilmatine, Mohamed. 1996. Un parler berbère-songhay du sud-ouest algérien (Tabelbala): Elements d'histoire et de linguistique. *Etudes et Documents Berbères* 14. 163–198.
- Trudgill, Peter. 2009. Sociolinguistic typology and complexification. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press.
- Van Coetsem, Frans. 1988. *Loan Phonology and the Two Transfer Types in Language Contact* (Publications in Language Sciences 27). Dordrecht: Foris.
- Wray, Alison & George W. Grace. 2007. The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* (The Evolution of Language) 117(3). 543–578. doi:10.1016/j.lingua.2005.05.005.
- Yanabu, Akira. 2011. *Kare* and *kanojo* - the shifting referents of two translation pronouns. In Indra Levy (ed.), trans. André Haag, *Translation in Modern Japan* (Routledge Contemporary Japan Series), 61-72. Abingdon: Routledge.