

Dmitry Idiatov

# An areal typology of clause-final negation in Africa



Language dynamics in space and time

**Abstract:** Clause-final negation markers (CFNMs), although typologically rare, can be found in a very wide range of languages of Northern Sub-Saharan Africa. Based on a sample of 618 African languages, this paper provides an analysis of spatio-temporal language dynamics in Sub-Saharan Africa with respect to the feature CFNM. I argue that it is important to consider together both the languages that have the feature under investigation and the languages that do not have it. Furthermore, in order to better capture the diversity of the languages that have CFNMs, I increase the degree of granularity of my data by taking into account two parameters, viz., obligatoriness of CFNMs and possible restrictions on the freedom to use CFNMs in different constructions. For spatial analysis and visualization, I use the methods of spatial interpolation and generalized additive modeling. Both methods converge on the need to distinguish two focal areas of the feature CFNM. The first one, the Central Focal Area, is the most prominent of the two and spans the east of West Africa and parts of Central Africa. The second one, the Western Focal Area, is less prominent and is restricted to West Africa. The two focal areas are separated by a major discontinuity around Ghana, Togo and Benin. In order to better calibrate the results of the spatial analysis and to identify the historical core of the Central Focal Area, I call onto other types of data available. Finally, I address the distribution of optional and/or restricted CFNMs in Africa, with a particular focus on the spread of CFNMs among Bantu languages to the south of the Central Focal Area, primarily in the Congo River corridor and the north of the Democratic Republic of Congo.

**Keywords:** African languages, areal typology, clause-final negation, historical linguistics, language contact, morphology, spatial analysis, syntax

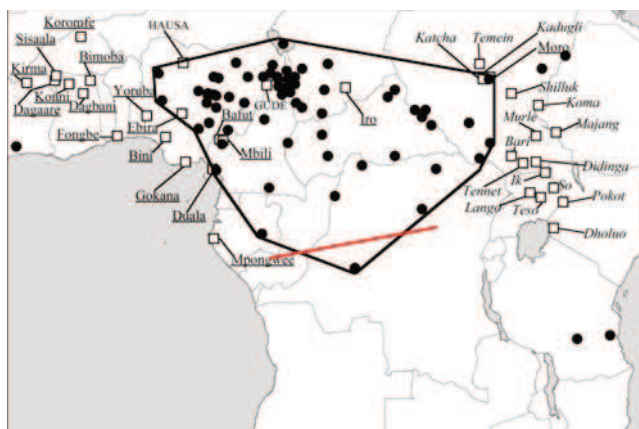
---

**Dmitry Idiatov**, LLACAN (CNRS – USPC/INALCO), 7, rue Guy Môquet BP8, 94801 Villejuif Cedex, France, [dmitry.idiatov@cnrs.fr](mailto:dmitry.idiatov@cnrs.fr)

 Open Access. © 2018 Dmitry Idiatov, published by De Gruyter.  This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. <https://doi.org/10.1515/9783110607963-005>

# 1 Introduction

In recent years, a number of studies, such as Beyer (2009), Dryer (2009) and Devos and van der Auwera (2013), have drawn attention to some typologically striking properties of negation marking in languages of different parts of Africa. Dryer (2009) focuses on “neutral negation”, i.e., obligatory and productive (general) negation marking patterns in declarative verbal main clauses expressed by negation markers that are words, in languages with SVO order in Africa. He demonstrates that SVO languages in “an area in central Africa [stretching] from Nigeria across to the Central African Republic and down into the northern Democratic Republic of Congo”, as illustrated in Figure 1,<sup>1</sup> significantly differ from SVO languages elsewhere in the world in that “the negative [word] follows the verb [instead of preceding it], typically occurring at the end of the clause, in SVONeg order” (Dryer 2009: 307). Dryer (2009) also points out that double negation marking is widespread in this region.

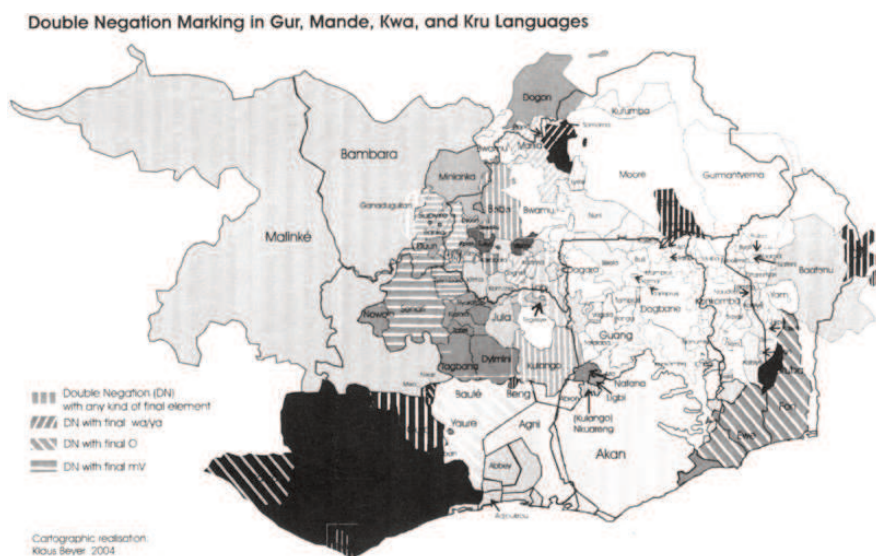


**Fig. 1:** VO&VNeg languages in Africa, with their core area delineated (Dryer 2009: 323)

Beyer (2009), in the same volume on negation patterns in West African languages (Cyffer, Ebermann and Ziegelmeyer 2009) as Dryer (2009), focuses specifically on

<sup>1</sup> The (red) line cutting off the southern-most point on the map has been added by me, as it must be a mistake. This point is labeled as Ngbaka (Ubangian) on other maps in Dryer (2009) but it has to be some Bantu language. Yet, no Bantu language from this area is mentioned in the paper.

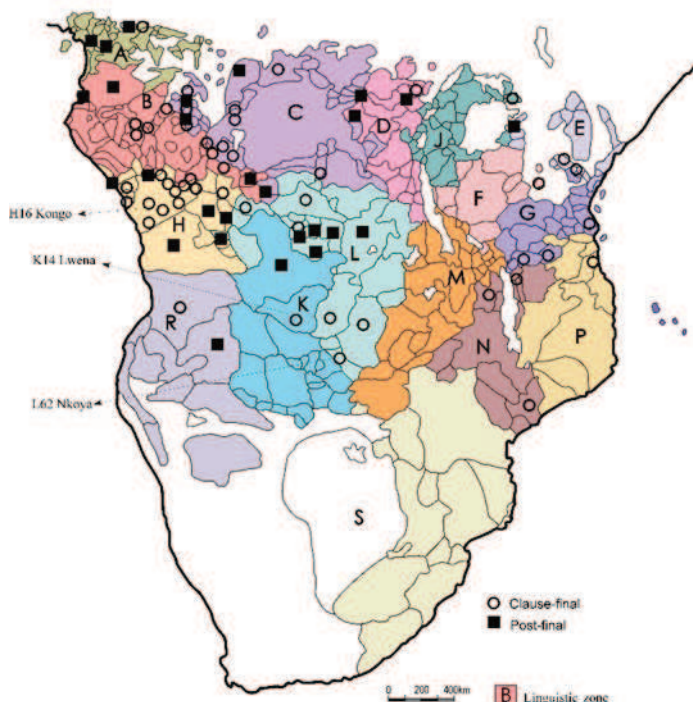
double negation marking for “sentential negation” in a large group of West African languages centered on the Volta River basin, as illustrated in Figure 2. In most cases, the second of the two negation markers also happens to be clause-final.



**Fig. 2:** Double negation marking in West African languages centered on the Volta River basin (Beyer 2009: 222)

Devos and van der Auwera (2013) is an in-depth study of multiple negation exponence in Bantu languages, a large group of languages spoken in a vast area from Cameroon and Kenya in the north all the way down to the South African Republic. They survey cases of multiple negation in Bantu languages, which is usually double but some exuberant examples of triple and quadruple negation marking are also attested. They also investigate recurrent sources for post-verbal negation markers. Many of these post-verbal negation markers happen to be also clause-final, as illustrated in Figure 3.<sup>2</sup>

<sup>2</sup> In the Bantuist tradition, the term “post-final” used in Figure 3 refers to the position immediately following the verb.

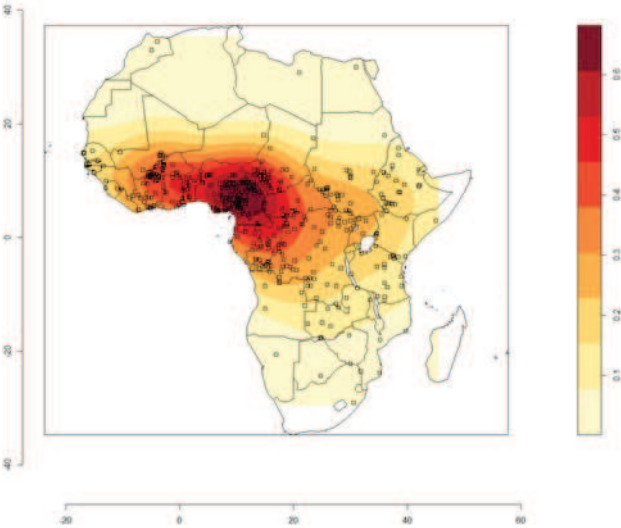


**Fig. 3:** Bantu double negation (Devos and van der Auwera 2013: 215)

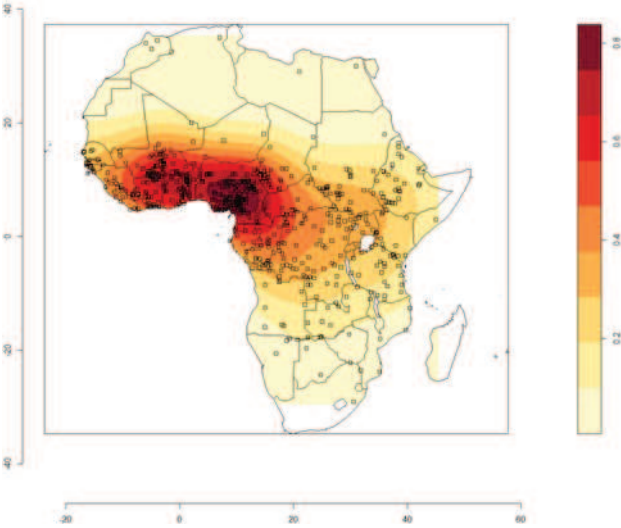
As these studies make clear, clause-final negation markers (CFNMs), although typologically rare, can be found in a very wide range of languages of Sub-Saharan Africa. Based on a sample of 618 African languages, I demonstrate in this paper that the spatial distribution of languages with CFNMs forms a clear areal pattern within Sub-Saharan Africa. At the same time, the spatial distribution of 462 languages with post-verbal negation markers of any kind does not form any distinctive areal pattern, as it is virtually identical to the spatial distribution of all the languages of the sample as a whole. The two distributions overlaid with their spatial intensity plots are shown in Figure 4 and Figure 5 respectively.<sup>3</sup> I am not able to plot the spatial distribution of multiple negation exponence on the scale of the

<sup>3</sup> All the plots and calculations for this paper have been produced with the software *R* (R Core Team 2015). The plots of spatial intensity and spatial interpolation have been produced with the package *spatstat* (Baddeley and Turner 2005).

continent at this point but I expect it to have a much less pronounced spatial structure than that of CFNMs.



**Fig. 4:** Geographic distribution of the 462 languages of the sample with post-verbal negation markers and a plot of their spatial intensity



**Fig. 5:** Geographic distribution of the 618 sample languages and a plot of their spatial intensity

Beside forming a clear areal pattern within Sub-Saharan Africa, on a world-wide scale CFNMs are also typologically much more unusual than post-verbal negation markers and multiple negation exponence. Furthermore, as I argued elsewhere (Idiatov 2012a), CFNMs in Sub-Saharan Africa tend to be characterized by a number of peculiarities in their morphosyntax and diachronic development that set them apart from similar markers elsewhere in the world and offer important clues as to an explanation of their observed areal distribution. Of course, some of these differences are more a matter of degree, yet some do seem to be more fundamental. For instance, CFNMs in African languages are often associated with the presence of multiple negation exponence within a clause, most commonly double but sometimes also triple and occasionally quadruple. CFNMs in Africa often happen to be morphosyntactically deficient as compared to more canonical grammatical markers in being optional or lacking in some types of clauses as conditioned by the TAM value of the predicate of the clause, the subordination status of the clause, the associated information structural and speech act type values or the discourse type that the clause belongs to (cf. Idiatov 2015). Diachronically, CFNMs in the area tend to be rather unstable and appear to be relatively easily borrowable (cf. Idiatov 2012b; 2015), unlike negators in other parts of the world but more like discourse markers, focus particles and phasal adverbs (cf. Matras 2009).

All this makes CFNMs in Sub-Saharan Africa a particularly interesting morphosyntactic feature to explore from the perspective of language dynamics in space and time. This paper provides such a spatio-temporal analysis of the feature CFNM in African languages. For reasons of space, I do not elaborate in the rest of paper on the explanation of why many negation markers are clause-final in Sub-Saharan Africa and why such negation markers are so common in particularly this region. I treat these issues in more detail elsewhere (Idiatov 2012a; in prep.). So here is just the gist of the explanation, which goes as follows. The clause-final position of the negation markers is explained by their origin in other clause-final markers. The fact that CFNMs are so common in this region is related to another typological feature of many of the relevant languages, viz., a grammatical category of clause-final markers whose core function is the expression of intersubjective meanings. Combined with the fact that negation is exactly one of those situations, propitious for the use of intersubjective markers, where the speaker's assertive authority is at stake, frequency effects account naturally for the tendency to conventionalize clause-final negation markers.

The paper is organized as follows. I begin by discussing in Section 22 various aspects of the definition of CFNMs adopted for this typology. This definition is rather inclusive because, as I explain in Section 2.1, my goal is to capture the most

of the synchronic diversity to achieve a better explanatory adequacy. However, for the reasons explained in Section 2.2, I leave negation constructions with nominal predicates out. Since I consider both obligatory and optional CFNMs, in Section 2.3, I address some of the issues that the distinction between the two may present. In Section 2.4, I elaborate on the meaning of the term “clause-final” within this typology. Related to the latter point is the issue of relevance of the relative order of object and verb for a typology of CFNMs. As I explain in Section 2.5, this order is not relevant in the typology presented here, unlike for example in the typology of post-verbal negation markers by Dryer (2009), which is confined to languages with VO order. I briefly present my sample in Section 3. I also provide maps of the 618 languages of the sample as a whole and of the languages with and without CFNMs. The simple binary division into languages that have and languages that do not have CFNMs hides important diversity among the languages with CFNMs. As I discuss in Section 4, in order to better capture this diversity and thus get a better idea of the possible historical and spatial dynamics behind the observed pattern, I increase the degree of granularity of my data by taking into account two parameters, viz., obligatoriness of CFNMs and possible restrictions on the freedom to use CFNMs in different constructions. Section 5 provides a discussion of the spatial and temporal dynamics reflected in the observed areal typological patterns of CFNMs in Africa. I first discuss in Section 5.1 the results and potential pitfalls of two methods of spatial analysis and visualization of the distribution of the values of the feature CFNM in Africa, viz., spatial interpolation and generalized additive modeling (GAM). Both methods converge on the need to distinguish two focal areas of the feature CFNM. The first one, the Central Focal Area (CFA), is the most prominent of the two and spans the east of West Africa and parts of Central Africa, largely coinciding with Dryer’s (2009) core area of VO&VNeg languages reproduced in Figure 1. The second one, the Western Focal Area (WFA), is less prominent and is restricted to West Africa. The two focal areas are separated by a major discontinuity around Ghana, Togo and Benin. In Section 5.2, I call onto other types of data to better calibrate the results of the spatial analysis produced in Section 5.1 and to identify the historical core of the CFA. Finally, Section 5.3 addresses the distribution of optional and/or restricted CFNMs in Africa, with a particular focus on the spread of CFNMs among Bantu languages to the south of the CFA, primarily in the Congo River corridor and the north of the Democratic Republic of Congo.

## 2 What kind of CFNMs are we looking at?

### 2.1 An inclusive definition: synchronic diversity as a window on language change

Morphosyntactic properties of negation constructions differ across languages. Similarly, the marking of negation may vary within a given language from one predicative construction to another. There are many different parameters along which the variation occurs. Depending on our goals and means, we can cut up this variation space in different ways. The definition that I adopt here is rather inclusive since my goal is to capture the most of the diversity. The rationale behind this is that synchronic diversity directly reflects the gradual nature of language change and thus offers us a window on the historical processes that brought about the current situation. Thus, one of the most common mechanisms known to be involved in the evolution of negation constructions, the so-called Jespersen cycle (cf. van der Auwera 2009, 2010 for a general overview; Devos and van der Auwera 2013 on Bantu languages), proceeds through a number of stages with most intermediate stages characterized by variation in the marking of negation within a given construction. Typically, related languages do not proceed on this path in exactly the same manner. Both the synchronic variation within one language and the synchronic diversity of negation patterns within a group of related languages offer an invaluable source of information on the earlier stages of the respective languages and on the processes underlying the change in negation constructions.

For the purposes of the present study, I consider as CFNMs the elements that may be used in the right periphery of negative verbal predications with clause scope negation but that do not appear in the corresponding positive predications and whose position is determined with respect to the clause as a whole. A CFNM may be the sole marker of negation in the clause or just one of the exponents of negation marking distributed within the clause. That is, my typology is not confined to double negation-marking, like the typology of Beyer (2009). A CFNM may be a dedicated negation marker or may also encode other meanings, such as tense, aspect, mood and emphasis (in this respect, see also Section 2.3). For the purposes of my typology, the degree of morphological bounding of CFNMs is not relevant either. That is, they may be words (similarly to the negation markers in Dryer's 2009 typology), clitics, (supra)segmental affixes or non-linear morphological operations. Similarly, my typology takes into consideration negation of all types of verbal clauses and is not restricted to the negation of declarative verbal main clauses, such as the "standard negation" typology of Miestamo (2008) or



the “neutral clausal negatives” typology of Dryer (2009). Negation of nominal predicates is beyond the scope of my typology for the reasons laid out in Section 2.2. As discussed in Section 2.3, I consider both obligatory and optional CFNMs. The meaning of the description “clause-final” in CFNM is explained in more detail in Section 2.4. Section 2.5 further elaborates on the clause finality of negation markers from the perspective of different relative orders of object and verb.

## 2.2 Beyond the scope of the typology: negation of nominal predicates

I am not concerned here with negation constructions with nominal predicates. The reason is not that I do not deem them relevant. Clearly, it is important to equally take into account negation strategies used with nominal predicates if one wants to achieve a comprehensive diachronic account of clausal negation constructions with verbal predicates. Thus, as pointed out by Croft (1991), negative existential markers may come to be extended to negative verbal predications within the so-called “negative-existential cycle”. However, from the perspective of an areal typology of CFNMs, negation constructions with nominal predicates tend to present rather different types of analytic problems. For instance, in the case of negative existential constructions (as distinguished from locative-presentative ones; cf. Veselinova 2013) that use a dedicated negation marker without any distinct existential marker, the question of whether the position of this marker is determined with respect to the clause as a whole or the nominal predicate may be simply irrelevant. For equational and identification constructions, it may not always be obvious which nominal should be considered the predicate (cf. Bisang and Sonaiya 2000 on Yoruba constructions of the structure X ‘BE’ Y, where X and Y are nominals). In view of these complications, another reason why I decided not to include negation of nominal predicates in the areal typology of CFNMs here is that it is my strong impression that their inclusion would not affect significantly the areal pattern established solely on the basis of constructions with verbal predicates. Thus, I found only a few languages described with CFNMs only in negation constructions with nominal predicates but not in the ones with verbal predicates, such as Ngangela [nba] (Bantu K12; Maniacky 2003), which has an optional CFNM *ko* only with nominal predicates as illustrated in (1) versus (2), and Beiya [kmy] (Adamawa, Samba Duru; Littig and Kleinewillinghöfer 2012), which has a CFNM <sup>2</sup>*wá* only with nominal predicates as illustrated in (3) versus (4) (for both languages, the sources provide only examples of identificational constructions). Although I do not consider these languages as having CFNM for the purposes of the typology presented here, their addition would not disrupt the

general areal pattern established on the basis of negation constructions with verbal predicates only.

(1) Ngangela<sup>4</sup>

- a. Kací                      *impweevó*      ko  
      NEG.COP              woman           NEG
- b. Kaci                      *ímpweevo*  
      NEG.COP              woman
- c. kéci-ko              *ímpweevo*  
      NEG.COP-NEG      woman  
      ‘It is not a woman.’  
      (Maniacky 2003: 192)

- (2) ko-*tw-a-mween-e*              *ḍingóómbe*  
      NEG-1PL-PRF-see.PR-NEG      cows  
      ‘We have not seen the cows.’  
      (Maniacky 2003: 140)

- (3) Beiya  
      yóó      yēn      kúsén      ʔwá  
      COP      thing      bush      NEG  
      ‘It is not a wild animal.’  
      (Littig and Kleinewillinghöfer 2012: 6)

- (4) Mìn                      túúrǎ                      Fàlélé  
      1SG                      come\NEG              PROP  
      ‘I do not come to Poli.’  
      (Littig and Kleinewillinghöfer 2012: 6)

---

<sup>4</sup> The following abbreviations will be used here: 1,2,3 first, second and third person; COP copula; DAT dative; DEF definite; DEM demonstrative; EMPH emphatic; FUT future; INDF indefinite; IPFV imperfective; LOG logophoric; NEG negation; NONHUM non-human; OBJ object; PFV perfective; PL plural; POSS possessive; PQ polar question; PRF perfect; PROG progressive; PROP proper name; PST past; QUO quotative; REFL reflexive; REL relative; SBJ subject; SBJV subjunctive; SG singular; STAT stative.

## 2.3 The issue of optionality

I consider both obligatory and optional CFNMs. Obligatory CFNMs may be obligatory throughout negation constructions or be confined to a subset of those. Optional elements of negation constructions are taken into consideration in so far as their addition does not change the propositional meaning of the negative predication or the constraints on their use are conditioned primarily by structural properties of their environment rather than their meaning (cf. Idiatov 2015 on the CFNM *wāā* in the Mande language Dzuun [dnn]). Admittedly, it is not always possible to make a clear-cut distinction along these lines precisely because language change is gradual. One of the frequent cases like this is represented by elements that are said to be optionally added to “emphasize” negation and are sometimes provided with translations such as ‘at all’. As a rule of thumb, I presume that if the author of a grammatical description deems it necessary to state that a negation construction may contain a given optional element, this element is frequent enough in this construction for its original referential meaning to be sufficiently backgrounded.

A different type of situation that is often conceived as involving optionality is when a default negation marker can be replaced by a negation marker that does change the propositional meaning of the negative predication and, for that reason, neither marker can be said to be obligatory as such, yet the presence of at least some such marker in the construction is required for the construction to be negative. In other words, it is the particular way of expressing negation within a negation construction that is obligatory but not the specific negation markers. French provides a good example of such a situation as a consequence of an ongoing Jespersen cycle type evolution and loss of negative concord (cf. van der Auwera and Van Alsenoy 2016). In colloquial French, the older preverbal negation marker *ne* is usually omitted and only the newer negation marker *pas* is used immediately following the verb, as in (5). The default negation marker *pas* can be replaced by a number of more specific markers, such as *jamais* ‘(n)ever’, as in (6), or *nulle part* ‘nowhere’, as in (7), and, although the latter elements do change the propositional meaning of the predication, at least some such element must be used in this constructional slot for the predication to remain negative. The alternative English translations of (6) and (7) using *never* and *nowhere* respectively are closer to the French original and have similar origins as well (cf. Ingham 2013).

### (5) French

*Elle* (ne) *va* *pas*.  
 She NEG goes NEG  
 ‘She doesn’t go.’

- (6) *Elle* (ne) *va* *jamais*.  
 She NEG goes never  
 ‘She doesn’t ever go.’ or ‘She never goes.’
- (7) *Elle* (ne) *va nulle part*  
 She NEG goes nowhere  
 ‘She doesn’t go anywhere.’ or ‘She goes nowhere.’

A somewhat more complicated example is provided by the Mande language Dzuun [dnn], as discussed by Idiatov (2015). Thus, Dzuun has a default CFNM *wāā*, as in (8), which may be omitted under certain conditions. In addition, Dzuun has a number of CFNMs that are semantically narrower than the default CFNM *wāā*, such as *dē* ‘anymore, no more’ and *kūrāā* ‘(n)ever; (not) at all’. These specific CFNMs usually stand alone, as in (9), replacing *wāā* just like *jamais* or *nulle part* replace *pas* in the French examples (6) and (7). However, occasionally, they can also be followed by *wāā*, as in (10), or they can co-occur with each other when the negative meaning needs to be further specified, as in (11). Finally, some of the forms that function as specific CFNM markers can also occur in positive constructions, as illustrated with *dē* in (12), where it functions as an emphatic marker. In this respect, consider French *jamais*, which can also be used in positive constructions, such as *si jamais* ‘if ever’ and *pour jamais* ‘forever’.

- (8) Dzuun  
 À náà wù è tsí wāā  
 3SG NEG.PST good 3SG.SBJV save NEG  
 ‘It was not good that he be saved.’  
 (Solomiac 2007: 270)
- (9) Wó dòn náà, wó nā bómà jàà dē  
 2SG enter come.IPFV 2SG NEG exit see.IPFV anymore  
 ‘You enter, but you do not find the exit anymore.’  
 (Solomiac 2007: 254)
- (10) Tà bwèy, bós rèè náà n’á rē yè ē  
 DEM moment Elder PL NEG.PST COP-3SG at 3PL.SBJV REFL  
 séré kúrāā wāā  
 pray at.all NEG  
 ‘At that time, the elders did not want to pray at all.’  
 (Solomiac 2007: 256, 578)

- (11) *Â náà fyā fyē dē kūrāā*  
 3SG NEG.PST fabric white anymore at.all  
 ‘[When the chicken wanted to come with the white fabric,] it was not a white fabric anymore.’  
 (Solomiac 2007: 539)
- (12) *Â cī, á! cī mún dzūnwēĩnsiá mún sàñ firū dē*  
 3SG QUO ah! QUO 1SG friend.DEF 1SG foot cheat.PFV EMPH  
 ‘He said: “Ah! My friend has really cheated me.”’  
 (Solomiac 2007: 483)

As illustrated on the example of Dzuun, a marker need not be a dedicated negation marker (be intrinsically negative in its meaning) to be considered a CFNM.

## 2.4 The meaning of being clause-final

The description “clause-final” in CFNM refers to the canonical position of the negation marker on the extreme right periphery of a clause. A given negation marker need not be in the absolute clause-final position in every possible construction to count as a CFNM. What is relevant is that, in the clause where the verbal predicate is accompanied by two or more simple nominal arguments and one simple adjunct modifying the predicate, such as a simple place or time adverbial, the position of the negation marker is determined with respect to the clause as a whole and not with respect to the verbal predicate, its nominal arguments or its modifier. In a given language, the position of the CFNM with respect to other right periphery markers and verbal predicate modifiers may be fixed or depend on a range of factors, such as their scope, meaning, morphosyntactic structure and length. Again, as in the discussion of optionality of CFNMs, a clear-cut distinction along these lines may not be always possible because change is gradual (although, more often, the difficulty is caused by the lack of relevant examples in the sources).

A good example of possible complexities involved in the syntax of CFNMs is provided by three Eastern Mande languages of the Boko-Busa cluster, Boko [bq̣c], Busa [bqp] and Bokobaru [bus], whose CFNMs have the form =o (Boko) and =ro (Busa and Bokobaru). Like all Mande languages, the languages of the Boko-Busa

cluster have a strict SOVX constituent order in transitive constructions<sup>5</sup> and SVX in intransitive constructions, where X stands for “oblique”, which is any constituent (an argument or an adjunct) other than S and O (cf. Creissels 2005). The canonical position of the negation marker  $=(\textit{r})\textit{o}$  is clause-final, as illustrated in (13). However, other right periphery elements with clausal scope, such as the polar question marker  $=\textit{\grave{a}}$ , follow the CFNM  $=(\textit{r})\textit{o}$ , as illustrated in (14). Furthermore, “sentence level adverbial phrases and clauses may follow the negative marker” (Jones 1998: 299), as illustrated in (15), which can be compared to (14). The tendency for adverbials to follow the negation marker  $=(\textit{r})\textit{o}$  is more general in Busa and Bokobaru while, in Boko, it is especially longer adverbials that are affected. Finally, the negation marker  $=(\textit{r})\textit{o}$  can be followed by the second coordinate in the alternative coordination construction, as in (16). This may be analyzed as a result of ellipsis, as is done by Jones (1998: 298). Alternatively, it may be seen as extraposition of a constituent to the right periphery because heavy constituents, such as the ones involving coordination, are dispreferred in argument positions (subject, object, postpositional phrase). This would not actually be uncommon in Mande and it would also parallel the tendency to place longer adverbials after the CFNM  $=(\textit{r})\textit{o}$ .

## (13) Boko

<i>ʔi</i>	<i>ĩ</i>	<i>gbě́</i>	<i>pĩ-ɔ</i>	<i>kã</i>	<i>lá</i>	<i>álé</i>
fluid	NEG.PFV	person	that-PL	intoxicate	as	2PL.PROG
<i>ʔe</i>	<i>wà=o</i>					
see	like=NEG					

‘Drink has not intoxicated those people as you are thinking.’  
(Jones 1998: 301)

(14) *ʔàsí* *álé* *ma* *náái* *kɛ* *ʔe* *tàa=o=à?*

so	2PL.PROG	1SG.POSS	trust	make	until	now=NEG=PQ
----	----------	----------	-------	------	-------	------------

‘So you are still not trusting me?’ (lit. ‘So you are not making my trust until now.’)  
(Jones 1998: 299)

5 Unlike most other Mande languages, the languages of the Boko-Busa cluster also allow null objects with anaphoric reading but only when the referent is non-human and only in non-perfective constructions, as well as perfective constructions with nominal subjects or a third person plural pronominal subject (Jones 1998: 212–213).

- (15) *aa mɛ̃ wa ʔi mi=o ʔe gɔɔ pɔ*  
 3PL.PFV say.PFV 3PL.LOG.FUT water drink=NEG until time REL  
*wà àà dɛ̃*  
 3.INDF.PFV 3SG.OBJ kill.PFV  
 ‘They said they would not drink until the time when he was killed.’  
 (Jones 1998: 299)

- (16) *má ʔésé vĩ=o ge màsé*  
 1SG.STAT sorghum have=NEG or maize  
 ‘I don’t have any sorghum or maize.’  
 (Jones 1998: 299)

The canonical position of the negation marker *=(r)o* in Boko-Busa is clause-final and this is how it is classified within this typology. At the same time, the observed synchronic variation in its placement is indicative of an ongoing diachronic process of the negation marker being attracted to the immediately post-verbal slot.<sup>6</sup>

## 2.5 CFNMs and the relative order of object and verb

Unlike in the typology of post-verbal negation markers by Dryer (2009), which is confined to languages with VO order, the relative order of object and verb is not relevant in the typology of CFNMs presented here. The object can either precede the verb as in the Dzuun and Boko-Busa examples above or follow it, as in the Gbaya Kara [gya] (Gbaya-Manza-Ngbaka) example in (17).

<sup>6</sup> The attraction of the negation marker in Boko-Busa from its original clause-final slot toward the immediately post-verbal one is likely to have been triggered by substrate influence of Baatonum, a Gur language spoken immediately to the southwest of Boko-Busa. In Baatonum, negation markers are mostly preverbal, except in the negative perfective construction, where the preverbal negation marker is complemented by a verbal suffix (Winkelmann and Miede 2009: 181–182). In Mande, the placement of negation markers varies but CFNMs are never attracted to the immediately post-verbal position, as it is not the position associated with polarity marking in Mande languages. Furthermore, there are sufficient reasons to assume that a substantial part of the current Boko-Busa populations shifted to Boko-Busa from Baatonum at some point in the past. For instance, Jones (1998: 5) points out the clear relation between the Boko-Busa terms for the non-royal Boko-Busa people (‘peasants’, ‘vassals’, ‘slaves’) and the Boko-Busa designations of the Baatonum.

## (17) Gbaya Kara

ʔám gbé sàdî há kòò kóm ɲóŋ ná  
 1SG kill\IPFV animal so.that wife POSS.1SG eat\IPFV NEG  
 ‘I did not kill game to feed my wife.’ (lit. ‘so that my wife eats’)  
 (Roulon-Doko 2012 : 5)

What is relevant for my typology is that the position of the negation marker on the right periphery is determined with respect to the clause as a whole. Constructions with VO order and constructions with OV order may present different types of analytic problems for determining whether the negation marker is clause-final in this sense or not.

As pointed out by Dryer (2009: 319), in those (Sub-Saharan) African languages with VO order where the negation marker follows the object it “predominantly” also follows “any adverbs or adjunct phrases”. In other words, it is typically a CFNM. In this respect, Sub-Saharan African languages differ from languages with VO order and the negation marker following the object elsewhere in the world, such as German, the language cited by Dryer (2009) as an example. A rare example of a language from Sub-Saharan Africa similar to German is Jur Mödö [bxé] (Bongo-Bagirmi; Andersen 1981; Persson and Persson 1991), spoken in South Sudan on the periphery of the core CFNM area (cf. Figure 8, 10 or 11). Jur Mödö uses SVX order in intransitive constructions and SVOX order in transitive constructions. The slot immediately at the end of the verb phrase, viz., after V in intransitive construction and after O in transitive construction or, framed differently, immediately before the X slot, appears to be reserved in Jur Mödö for at least two grammatical markers, one of which is the negation marker *dé*, as illustrated in (18) and (19), and the other one is the resultative or perfect marker *déní* (called “perfective” by Andersen 1981 or “completive” by Persson and Persson 1991, as illustrated in (20)).<sup>7</sup>

<sup>7</sup> Contrary to Dryer’s (2009: 320) statement that the negation marker “can be freely positioned among adverbial or adjunct elements”, Andersen (1981) and Persson and Persson (1991) only specify that the negation marker is a type of “adverb” which should “occur in the adjunct [position], separated from the verb by the object” (Persson and Persson 1991: 15). Yet, in all the examples found in these sources, the negation marker is always the first of the “adverbs” or “adjuncts”, immediately following the verb or, if present, the object.



## (18) Jur Mödö

*m-óḁḁ ndòbò dé kpè tí=ì*  
 1SG-do work NEG again with=2SG

‘I won’t work with you again.’

(Andersen 1981: 59)

(19) *mòró ìlábá dé rò kòbì*

spear fall NEG at buffalo

‘The spear did not hit the buffalo.’

(Andersen 1981: 80)

(20) *kirábà òpè kómó ǵení dī mī mālībīwù*

jackal release hare PRF from in snare

‘Jackal released Hare from the snare.’

(Persson and Persson 1991: 15)

Beside the typical situation in African VO languages, where the clause-final status of a negation marker is relatively straightforward, we also find a number of VO languages on the periphery of the core CFNM area (cf. Figure 10 or 11), where a negation marker gravitates towards the end of the clause but it is not obvious whether its canonical position should be characterized as clause-final or not. One of the clearest examples of such a language is Nzadi [no code] (Bantu B865; Crane, Hyman and Tukumu 2011), whose description provides a detailed overview of the syntax of the post-verbal negation marker. In Nzadi, the negation is marked in two positions in the clause with the first marker occurring before the verb “in the auxiliary” (the form of this negation marker depends on the TAM values) and the second marker, *bɔ*, occurring after the verb “towards the end of the clause” and taking scope “over any of the elements” of the clause (Crane, Hyman and Tukumu 2011: 169, 173). (21) schematizes the possible positions of *bɔ* in various clause structures.

- (21) Nzadi: the possible positions of the post-verbal negation marker *bɔ* (Crane, Hyman and Tukumu 2011: 171)<sup>8</sup>

S-V- <i>bɔ</i>	*S- <i>bɔ</i> -V	* <i>bɔ</i> -S-V
S-V-O- <i>bɔ</i>	?S-V- <i>bɔ</i> -O	
S-V-IO-DO- <i>bɔ</i>	S-V-IO- <i>bɔ</i> -DO	*?S-V- <i>bɔ</i> -IO-DO
S-V-DO-Obl- <i>bɔ</i>	?S-V-DO- <i>bɔ</i> -Obl	*?S-V- <i>bɔ</i> -DO-Obl
S-V-DO-Obl <sub>ben</sub> - <i>bɔ</i>	S-V-DO- <i>bɔ</i> -Obl <sub>ben</sub>	*S-V- <i>bɔ</i> -DO-Obl <sub>ben</sub>
S-V-X- <i>bɔ</i>	S-V- <i>bɔ</i> -X	

For the purposes of my typology, negation markers similar to Nzadi *bɔ* are classified as optionally clause-final. From a diachronic perspective, such indeterminacy suggests an ongoing syntactic change whereby a negation marker that, by virtue of its etymology, has originally evolved in a certain slot in the clause structure is being attracted to a different slot in the clause structure, presumably because this slot is associated with the expression of certain types of meanings.<sup>9</sup>

<sup>8</sup> The asterisk <\*> marks ungrammatical options. Elsewhere in the source, the examples of *bɔ* placement options marked with the combination <\*> are also characterized as “ungrammatical”, so it remains unclear what difference between <\*> and <?\*> was intended by the authors in this table. The question mark <?> marks options that are characterized as “strongly dispreferred” or “at least marginally acceptable”. *S-V-IO-DO* stands for ditransitive “double object constructions”, where the indirect object is unmarked. *S-V-DO-Obl* stands for ditransitive “indirect object constructions”, where the indirect object, referred to as oblique, is introduced by the locative preposition *kó*. I added to the original table the row with the benefactive oblique (*Obl<sub>ben</sub>*) marked by *sám* ‘é N (lit. ‘reason of N’), because it differs from the obliques introduced by the preposition *kó* in that “the preferred ordering may place *bɔ* before the benefactive”, although without any “strong preference either way” (Crane, Hyman and Tukumu 2011: 170). Finally, “X can be a non-object complement, or any adjunct, and may co-occur with direct and indirect objects [including obliques], with *bɔ* placement restricted with regard to objects as in other cases” (Crane, Hyman and Tukumu 2011: 171).

<sup>9</sup> The original position of the Nzadi marker *bɔ* is probably after the indirect object, either the unmarked one or the one introduced by the preposition *kó* or, in the absence of such an indirect object, after the direct object or, when no object is present, after the verb. That is, it is now being attracted to the clause-final position, arguably because of its default clausal scope and its intermediary function as attenuator of the assertive strength of the negative predication as a whole. Its original placement can be explained by its likely etymology as a possessive pronoun, which used to be coreferential with the subject and functioned as a kind of attenuator, something like ‘as for X<sub>i</sub> [S<sub>i</sub> does not P]’, which can be roughly compared to some uses of emphatic pronouns in French, as in *Pierre ne sait pas, lui* ‘Pierre does not know (while others might know)’ (lit. ‘Peter does not know, him’). Given the shape of *bɔ*, it is most likely the third person plural form that has become generalized. As described by Devos and van der Auwera (2013), possessive pronouns are not uncommon as a source of secondary negation markers in the Bantu languages of the area.

African languages with OV order and a post-verbal negation marker can be subdivided into two groups for the purposes of my typology. In the first group, the verb is normally followed by some constituents (arguments or adjuncts) other than the object. Most such languages seem to behave like the Mande languages Dzuun and Boko presented in Sections 2.3 and 2.4 in that the post-verbal negation marker also follows other post-verbal constituents and thus can be characterized as clause-final. Most such languages are in fact Mande. In the second group, the clause is basically verb-final so that, in principle, the question of whether the post-verbal negation marker is oriented toward the clause as a whole or just the verb is not particularly meaningful. However, where some diachronic evidence is available, it is usually clear that the post-verbal negation marker is oriented toward the verb and not the clause as it often originates in a main verb reanalyzed as an auxiliary (cf. van Gelderen 2008: 232–233; Lucas 2009 on Afro-Asiatic languages). Therefore, by default, the post-verbal negation markers in such languages are not characterized as clause-final for the purposes of my typology. This situation is common among the Afro-Asiatic languages of northern and eastern Africa (Cushitic, Omotic, Semitic), as illustrated in (22) from Dhasaanac [dsh] (Cushitic; Tosco 2001), in some Nilo-Saharan groups in Chad and Sudan (such as Saharan, Fur and Nubian) and in Dogon and Ijoid languages in western Africa, as illustrated in (23) from Jamsay [djm] (Dogon; Heath 2008).

(22) Dhasaanac

<i>yáa</i>	<i>ǰúm</i>	<i>ma</i>	<i>ká</i>	<i>šuggun-ij̣</i>
1SG.SBJ	children	NEG	here	bring.IPFV-NEG

‘I am not going to bring the children here.’  
(Tosco 2001: 299)

(23) Jamsay

<i>ýýró</i>	<i>kò-rú</i>	<i>yèwò-l-á</i>
quickly	NONHUM-DAT	accept-PFV.NEG-3PL.SBJ

‘They did not readily accept it [= plow].’  
(Heath 2008: 368)

### 3 The data

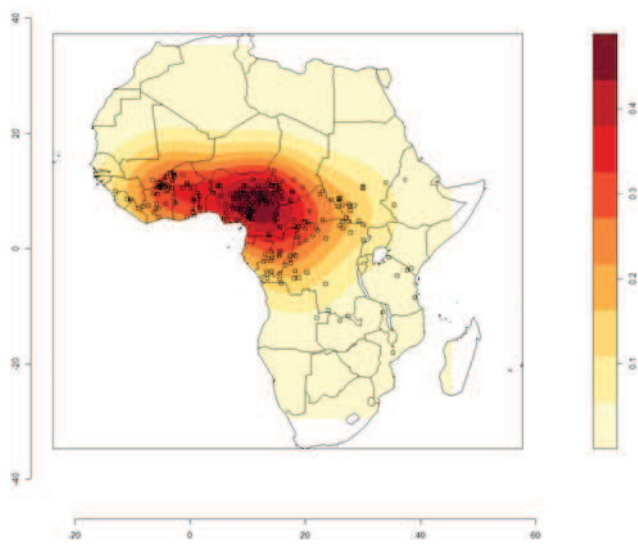
The data for this study come from individual grammatical descriptions complemented by a number of existing typological surveys of negation patterns in Af-

rica, such as Dryer's (2009) survey of post-verbal negation markers in the VO languages of Central Africa, Devos and van der Auwera's (2013) study of multiple negation marking in Bantu languages<sup>10</sup> and Beyer's (2009) study of double negation marking in the languages of the area centered around the Volta River basin in western Africa. I tried to cross-check the information coming from typological surveys in grammatical descriptions whenever possible.

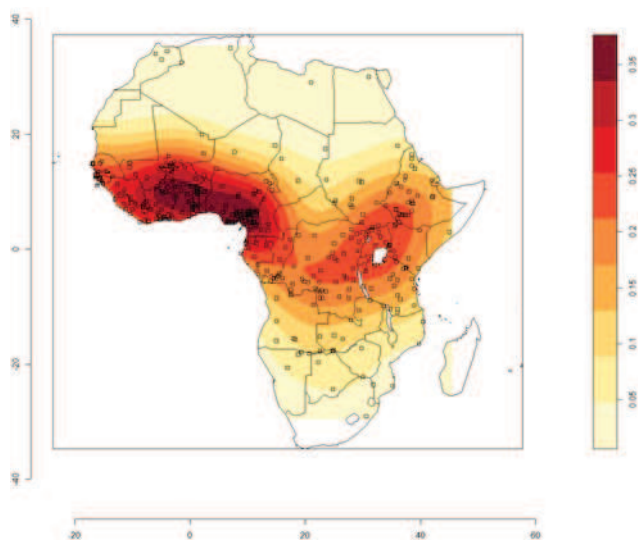
My sample consists of 618 languages, of which 256 languages appear to use some kind of CFNM while 328 languages clearly lack a CFNM and, for 34 languages, the information available was not sufficient for an informed decision. For most purposes, I combined the latter two groups as languages without CFNMs (362 languages). The geographic distribution of the 618 languages of my sample is presented in Figure 5 in Section 1. Figure 5 also represents this distribution as spatial intensity, that is, the degree of concentration of languages taken as points in space. The most important concentration of languages is found in the area around the border between Cameroon and Nigeria. Another area of high concentration of languages stretches from Togo into the southwest of Burkina Faso. Figure 6 shows the geographic distribution and the spatial intensity of the 256 languages that have CFNMs and Figure 7 of the 362 languages that do not have CFNMs. The overall pattern of distribution of languages with CFNMs in Figure 6 resembles the pattern of distribution in the sample as a whole in Figure 5. The pattern in Figure 6 is, however, more spatially circumscribed in almost all directions. It is basically restricted to northern Sub-Saharan Africa. Its focal area, although equally situated in the area around the border between Cameroon and Nigeria, has a relatively northern position and its westward extension towards southwestern Burkina Faso is somewhat less pronounced and has a more clearly latitudinal east-west orientation (as opposed to a more southeast-northwest orientation in Figure 5).

---

<sup>10</sup> I am grateful to Maud Devos and Johan van der Auwera for providing me with the source database they created for that survey.



**Fig. 6:** Geographic distribution of the 256 languages with CFNMs and their spatial intensity



**Fig. 7:** Geographic distribution of the 362 languages without CFNMs and their spatial intensity

The pattern of distribution of languages without CFNMs in Figure 7 is quite different, especially in its eastern part. To begin with, the distribution in Figure 7 is

much more spread out. Importantly, it is characterized by a clear depression in Central Africa where the pattern has a U-shaped curve. This depression in the pattern in Figure 7 is created by the presence of a relatively homogenous cluster of languages with CFNMs in that area, whose importance may not be obvious from Figure 6. The depression (and the cluster of languages with CFNMs that creates it) is northeast-southwest oriented and extends from the Central African Republic along the Congo River, corresponding on the map to the border between Congo and the Democratic Republic of Congo. This depression makes the pattern go southward in Cameroon, parallel to the coast towards the lower reaches of the Congo River, where it then turns eastward and finally turns back around central Democratic Republic of Congo in a northeast direction toward Ethiopia. It is interesting to compare the West African focal area of Figure 7 with that of Figure 6. The focal area in Figure 7 is both wider and more pronounced, stretching in a southeast-northwest orientation similar to what we find in the sample as a whole in Figure 5. Similarly to the focal area in Figure 6, the eastern end of the focal area in Figure 7 is situated around the border between Cameroon and Nigeria but it has a clearly more southern position, spanning southeastern Nigeria and southern Cameroon.

## **4 Increasing the granularity in the data: obligatoriness and constructional freedom**

The patterns of geographic distribution of the languages of the sample presented in Figure 5 (Section 1) and Figures 6 and 7 in Section 3 are basically point patterns. As such, they show us the overall extent of the languages with and without CFNMs and highlight the regions of high and low concentration of the two types of languages. This is valuable information for a first approach to the phenomenon. However, this binary representation hides important diversity among the languages with CFNMs. In order to better capture this diversity and thus get a better idea of the possible historical and spatial dynamics that brought about the current situation, we need to increase the degree of granularity of our data. Following the discussion of the different issues involved in delimiting CFNMs in Section 2, I will use two parameters. The first one is obligatoriness of CFNMs, that is, whether CFNMs are obligatory or optional, and the second one are possible restrictions on the freedom to use CFNMs in different constructions. The two ranking options of these parameters and the pseudo-numerical values assigned to

them are summarized in Table 1.<sup>11</sup> The last column provides the numbers of languages of each type. CFNMs that are obligatory and free from constructional restrictions are ranked highest, as 4, in both ranking options, while CFNMs that are both constructionally restricted and optional are ranked lowest, as 1.

**Tab. 1:** Constructional restrictions and optionality: two ranking options

Constructional freedom highest	Constructional freedom	Obligatoriness	Obligatoriness highest	Number of languages
0		no CFNMs	0	328
0.5		unclear	0.5	34
1	restricted	optional	1	7
2	unrestricted	optional	2	22
3	restricted	obligatory	3	31
4	unrestricted	obligatory	4	196

In principle, either of the two parameters could be ranked first. It so happens that, for this particular distribution of languages with CFNMs, which is highly skewed to one side, both options produce very similar results. Nevertheless, I have a principled preference for ranking obligatoriness highest because I conceive obligatoriness as the defining property of grammatical meanings (see Idiatov 2008 for a detailed discussion). Lack of constructional restrictions on the use of a grammatical marker is a property of canonical grammatical markers (in the sense of canonical typology; cf. Brown, Chumakina and Corbett 2013). Therefore, CFNMs that are both obligatory and free of constructional restrictions are canonical grammatical markers whereas other types of CFNMs fall short of such status.

An important point to be mentioned with respect to the classification in Table 1 is that it classifies languages, not CFNMs. If a language has several CFNMs that

<sup>11</sup> The values are pseudo-numerical in the sense that their numeric values are basically arbitrary and are just intended to reflect the relative order of the different combinations of the parameters. In fact, whether we use these numeric values or just an ordered list of factors does not matter that much. The two methods give very similar results in terms of spatial analysis, for instance, when we visualize them using spatial interpolation (see Section 5.1). However, I prefer to use the pseudo-numeric values because they allow to better capture the relative status of languages that do not have CFNMs as opposed to the different types of languages that have or may have CFNMs. See also Section 5.1 for an alternative coding scheme for pseudo-numeric values applied in generalized additive modeling.

differ with the respect to the two parameters, I choose the CFNM that ranks highest, as the closest to being a canonical grammatical marker, to represent the language as a whole. Admittedly, this way, some of the diversity fails to be properly reflected in the typology but I do not see very well how I can incorporate this information. Furthermore, I also have the impression that adding it would not have significant effects on the overall results.

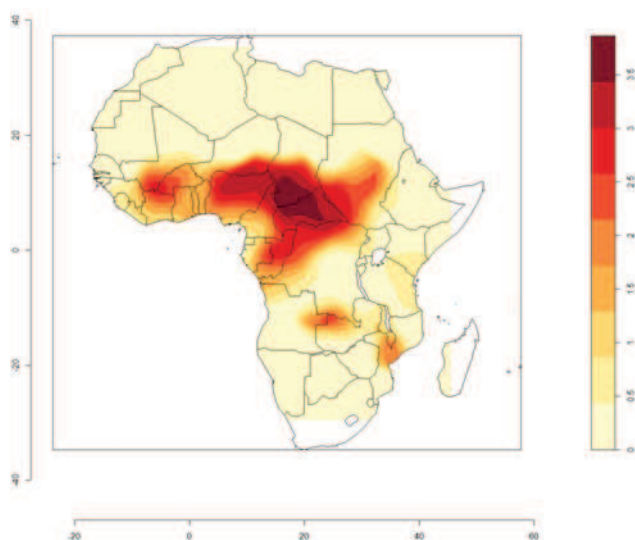
## 5 Areal typology of CFNM in Sub-Saharan Africa

In this section, I first discuss the results and potential pitfalls of two methods of spatial analysis and visualization of the distribution of the values of the feature CFNM in Sub-Saharan Africa, as well as some geographic correlations that emerge from this analysis (Section 5.1). In particular, I apply spatial interpolation (using two different types of smoothing, kernel smoothing and inverse-distance weighted smoothing) and generalized additive modeling (GAM). The different methods used converge on the same spatial pattern of the feature CFNM. They confirm the existence, the position and the overall shape of two focal areas, the Central Focal Area (CFA) spanning the east of West Africa and parts of Central Africa and the Western Focal Area (WFA) restricted to West Africa. The two areas are separated by a major discontinuity around Ghana, Togo and Benin. Of the two focal areas, the CFA can be called the primary focal area, given its prominence, and the WFA a secondary focal area. In Section 5.2, I address the issue of the historical core of the CFA. In particular, I argue that, despite the apparent prominence of an area in southern Chad and the Central African Republic within the CFA, it cannot represent its historical core and that it is much more likely that the primary historical core of the CFA is situated immediately to the northwest of the Central African Republic along the Benue River corridor going from southern Chad through northern Cameroon into central Nigeria. At the same time, as discussed in Section 5.3, this area in southern Chad and the Central African Republic prominent within the CFA must have served as the source for the spread of the feature CFNM among Bantu languages further south in the Congo River corridor and the north of the Democratic Republic of Congo. Section 5.3 further offers a discussion of the broader issue of the distribution of optional and/or restricted CFNMs in Africa and argues that, as expected, such grammatically non-canonical CFNMs tend to be peripheral areally as well.

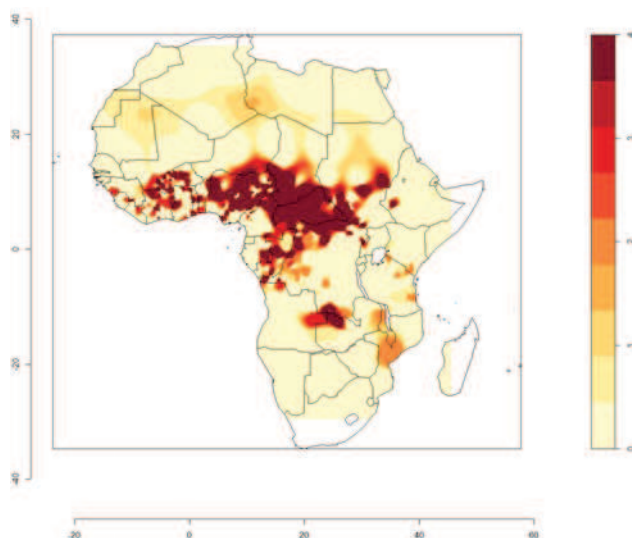


## 5.1 Spatial analysis: spatial interpolation, generalized additive modeling and correlations with geography

The spatial distribution of languages with different kinds of CFNMs (as distinguished in Section 4) and languages without CFNMs can be inspected in a number of ways. The most straightforward option is to visualize the data by means of spatial interpolation, which I perform here using the pseudo-numeric values described in Section 4. We can also use an alternative coding scheme for pseudo-numeric values, as we will do for generalized additive modeling further in this section, or use an ordered list of factors without any noticeable impact on the results. Thus, Figure 8 shows the result of spatial interpolation using kernel smoothing and Figure 9 shows the result of spatial interpolation using inverse-distance weighted smoothing.



**Fig. 8:** The spatial interpolation graphic of the different values of the feature CFNM (as described in Section 4) using Gaussian kernel smoothing (the default bandwidth value adjusted by 1.3)



**Fig. 9:** The spatial interpolation graphic of the different values of the feature CFNM (as described in Section 4) using inverse-distance weighted smoothing (power = 6)

Different spatial interpolation methods produce slightly different visualizations, which may allow to better highlight different aspects of the spatial distribution. Thus, the interpolation using kernel smoothing in Figure 8 is somewhat better in visualizing the overall structure of the spatial distribution of the feature CFNM. It clearly shows a major discontinuity in the distribution of languages with CFNMs in Northern Sub-Saharan Africa (NSSA) around Ghana, Togo and Benin. This discontinuity cuts off a secondary focal CFNM area that is centered on the region where the borders of Burkina Faso, Mali and Ivory Coast come together. For ease of reference, I refer to this secondary CFNM focal area in NSSA as the Western Focal Area (WFA) and to the main CFNM focal area to the east of it as the Central Focal Area (CFA). Both the interpolation using kernel smoothing in Figure 8 and the interpolation using inverse-distance weighted smoothing in Figure 9 show that the CFNM areas in NSSA are largely confined to the hinterland. The location of the three clearest extensions of the CFNM area toward the Gulf of Guinea coast is somewhat better visible in Figure 9. Thus, the first such extension is found around southern Togo and Benin, largely bridging the gap between the CFA and the WFA. The second coastal extension is located in the south of central Nigeria and is formed by the southward spread of the Edoid languages. While the first two coastal extensions are themselves likely to result from relatively recent language spread and/or contact events, the gap between them may be just as well

accidental. Thus, the gap area is occupied by one big language (or a cluster of closely related lects), viz., Yoruba, that must have expanded into the gap area from a more hinterland location in central Nigeria relatively recently and the respective proto-language may have simply happened to lack the CFNM feature by chance or lost it when moving into the area. In this respect, note that CFNMs are found in various related languages spoken just outside of the gap area, such as many Edoid languages or Igala, which belongs to the same lower-level Yoruboid linguistic grouping as Yoruba. The third coastal extension along the Congo River corridor is due to relatively recent language and/or population movements out of Central Africa affecting the Bantu languages in that area (see Section 5.3).

Depending on how the underlying data is distributed in space exactly, spatial interpolation may produce certain visualization artefacts that one should be aware of when analyzing the results. Thus, both methods exaggerate to different extents the prominence of a number of regions, such as the region where the borders of the Democratic Republic of Congo, Angola and Zambia come together, the region in central Mozambique and the region toward the northeast of South Sudan. These exaggerated prominence regions are due to the fact that the sample data points are sparsely distributed in these regions (either because there are simply fewer languages or because the languages were not sampled). To see why this may affect visualization, we can represent data points by peaks whose height corresponds to the numeric value of the types in Table 1. The peaks representing a few isolated examples of CFNM languages in such a region would get very wide slopes when other data points are far. Several low prominence regions in the Sahara in Figure 9 are basically due to the same reason and do not correspond to any real languages, as becomes clear when we compare Figure 9 to Figure 6. That these spurious prominence regions in the Sahara are absent in Figure 8 is just due to the way the sample data points happen to be distributed on the southern fringes of the Sahara and the bandwidth value chosen for kernel smoothing.<sup>12</sup> Finally, care should be exercised when interpreting the region of high prominence (as reflected by its darker shading) within the CFA in southern Chad and the Central African Republic in Figure 8. Although it is tempting to interpret it as the core or hotbed of the CFA, as discussed in Section 5.2, a different interpretation may be more appropriate. In this respect, note that this region of high prominence is absent in Figure 9, which uses a different spatial interpolation method.

---

<sup>12</sup> In this respect, note a number of slight northward spikes in Figure 8, such as the spikes in central Chad and southeastern Niger, which correspond to much clearer spikes in the same places in Figure 9.

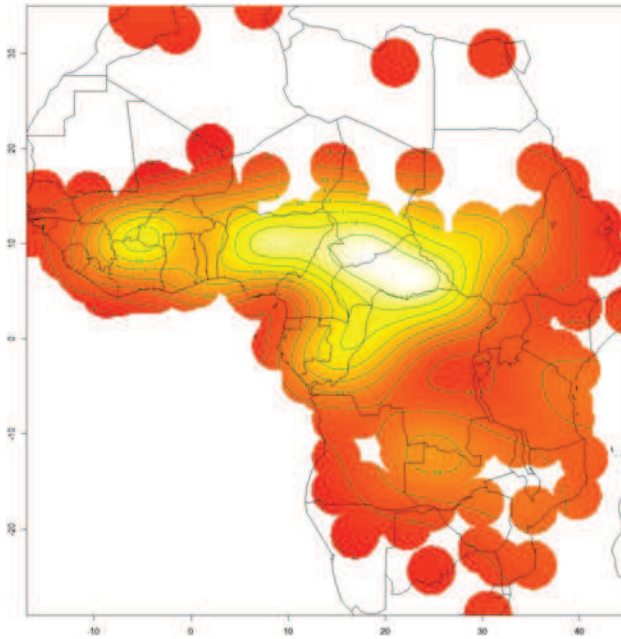
Although spatial interpolation is a valuable visualization tool, it does not produce quantifiable results. A statistical tool well-adapted for spatial analysis that can do that is provided by generalized additive models (GAM),<sup>13</sup> for instance, as implemented in the *mgcv* package for R (Wood 2015). There are different ways to code our response variable, viz., the type of the feature CFNM. One option would be to code it as an ordered categorical variable. The big disadvantage of this option would be that it does not reflect the importance of the divide between the absence and presence of CFNMs in the language and the certain hierarchy between the six values of the feature CFNM of Table 1. For this reason, I will not use this coding. The results it produces are actually largely comparable to the other two options that we are going to consider, although less clear-cut. The latter two options both involve pseudo-numeric values assigned to the six values of the feature CFNM. The first coding scheme simply reuses the numbers from Table 1 (with obligatoriness ranked highest). The other coding scheme uses a scale from 0 to 1, with 0 corresponding to the absence of CFNMs and 1 corresponding to the presence of canonical CFNMs (type 4 in Table 1), which may better capture the hierarchy between the six values of the feature CFNM. The two coding schemes are compared in Table 2.

**Tab. 2:** Two coding schemes for pseudo-numeric values of the feature CFNM

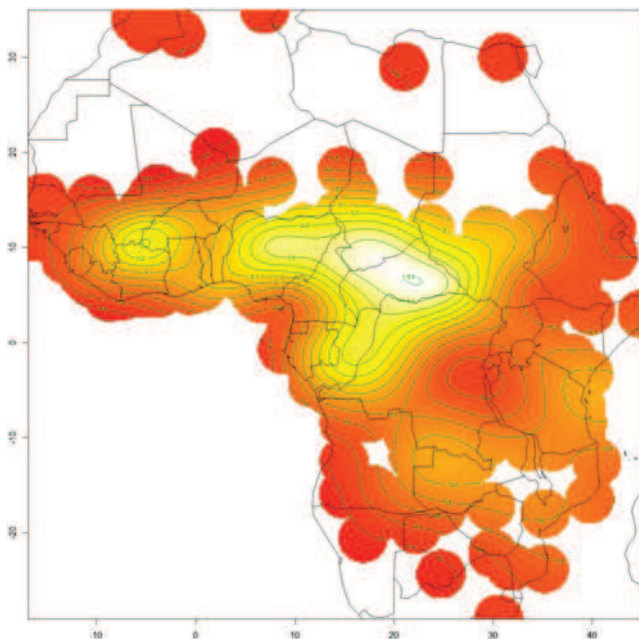
Constructional freedom	Obligatoriness (ranked highest)	Scheme 1	Scheme 2
	no CFNMs	0	0
	unclear	0.5	0.5
restricted	optional	1	0.625
unrestricted	optional	2	0.75
restricted	obligatory	3	0.875
unrestricted	obligatory	4	1

<sup>13</sup> GAM is an extension of multiple regression that provides flexible tools for modeling complex interactions describing wiggly surfaces. A practical introduction to GAMs for linguists can be found in Baayen (2013), Tammimga, Ahern and Ecay (2016) and Winter and Wieling (2016). Some good examples of use of GAMs in linguistics in relation with spatial analysis are provided by Wieling, Nerbonne and Baayen (2011) and Wieling et al. (2014). Idiatov and Van de Velde (2016, 2018) apply GAM for spatial analysis of lexical frequencies of labial-velar stops in NSSA.

We will consider two GAMs, one using Scheme 1 and one using Scheme 2. Both GAMs estimate the values of the feature CFNM as a function of the combination of longitude and latitude using thin-plate regression splines. The plots in Figure 10 and Figure 11 represent the regression surface of the two GAMs produced using Gaussian distribution as a contour plot with the heat map color scheme. In the heat map color scheme, the lighter the color, the higher the temperature, which, in our case, corresponds to a higher pseudo-numeric value of the feature CFNM. The contour lines are isopleths that mark deviations from the mean in terms of standard deviation.



**Fig. 10:** A contour plot with the heat map color scheme visualizing a GAM produced using Scheme 1 for coding of the feature CFNM ( $k=13$ , family = Gaussian, edf = 41.73,  $p < 2e-16$ , deviance explained = 43.7%, AIC = 2234)



**Fig. 11:** A contour plot with the heat map color scheme visualizing a GAM produced using Scheme 2 for coding of the feature CFNM ( $k=13$ , family = Gaussian,  $\text{edf} = 42.72$ ,  $p < 2e-16$ , deviance explained = 42.5%, AIC = 552)

The two GAMs produce very similar results. Basically, they only differ in their Akaike information criterion values (AIC), with the GAM based on Scheme 2 having a much better AIC. However, this does not so much reflect the quality of the respective model as the difference in the coding scheme used, arguably favoring Scheme 2.

The GAM plots in Figure 10 and Figure 11 are very similar to the spatial interpolation plots in Figure 8 and Figure 9, minus the visualization artefacts produced by spatial interpolation and discussed above. The different methods used converge on the same spatial pattern of the feature CFNM. They confirm the existence, the position and the overall shape of two focal areas, the CFA and the WFA, separated by a major discontinuity around Ghana, Togo and Benin.

Of the two focal areas, the CFA can be called the primary focal area, given its prominence, and the WFA a secondary focal area. The overall shape of the two focal areas can be broadly described as the hinterland of the Gulf of Guinea. Starting with the WFA in southern Mali and northern Ivory Coast, the region where the feature CFNM is prominently present extends eastward, primarily following

grasslands and woodland savannahs north of the forest zone, for the most part staying outside of the coastal regions. It is interrupted by only one major discontinuity separating the WFA from the CFA. Geographically, this major discontinuity corresponds rather well to the so-called Dahomey Gap, a southward savannah corridor interrupting the zonal West African rain forest. This may look strange at first, as the WFA and the CFA themselves are in the savannah zone. However, the north-south orientation of this savannah corridor can also be seen as conducive to interrupting the general east-west dynamics of the population and language movements in the western part of NSSA. Thus, I believe that this discontinuity is primarily due to the combined effect of the southward spread of Songhay into the current gap area from the north and the northward spread of the Tano subgroup of Kwa from the coastal regions in the south.

The slight southeastward bent of the CFA in southern Chad toward the Central African Republic follows well the orientation of the relevant ecological zones, topography and hydrography of this part of NSSA, as illustrated in Figure 12 on a relief map of Africa.<sup>14</sup> The CFA further marginally spills over into South Sudan in the east and, much more significantly, into equatorial Central Africa in the southwest along the Congo River corridor (see Section 5.3). The two plausible zones through which the interaction between the CFA and these neighboring areas is likely to have primarily occurred are marked in Figure 12 as A and B respectively, with the difference in the degree of interaction graphically represented by the difference in font style of the two symbols.

---

**14** The eastern borders of Chad and the Central African Republic largely correspond to the divide separating the Lake Chad and Congo River drainage areas to the west from the Nile River drainage area further east. The southwestern border of Chad and the western border of the Central African Republic roughly reflect the divide separating the same Lake Chad and Congo River drainage areas from the drainage areas of the Niger River and some smaller rivers flowing into the Gulf of Guinea. The shape of the divides themselves naturally results from the relief of the area.



**Fig. 12:** A relief map of Africa: the arrow highlights the correlation between the topography and the southeastward bent in the CFA around southern Chad – A and B mark plausible primary interaction zones between the CFA and the neighboring regions in South Sudan and equatorial Central Africa

## 5.2 The Central Focal Area: the core issue

In Figure 8, which is a spatial interpolation plot with kernel smoothing, and Figures 10 and 11, which are visualizations of two different GAMs, the CFA is characterized by a region of high prominence in southern Chad and the Central African Republic (as reflected by its darker shading in Figure 8 and lighter colour in Figures 10 and 11). Therefore, it may be tempting to interpret the region in southern Chad and the Central African Republic as the core or hotbed of the CFA, with all the obvious historical implications that such an interpretation would entail. At the same time, this region of high prominence is absent in Figure 9, which uses a different spatial interpolation method, viz., inverse-distance weighted smoothing. Similarly, the same region looks anything but prominent in Figure 6, which shows the distribution of the languages with CFNMs and their spatial intensity. Clearly, care should be exercised when interpreting the relevance of this region



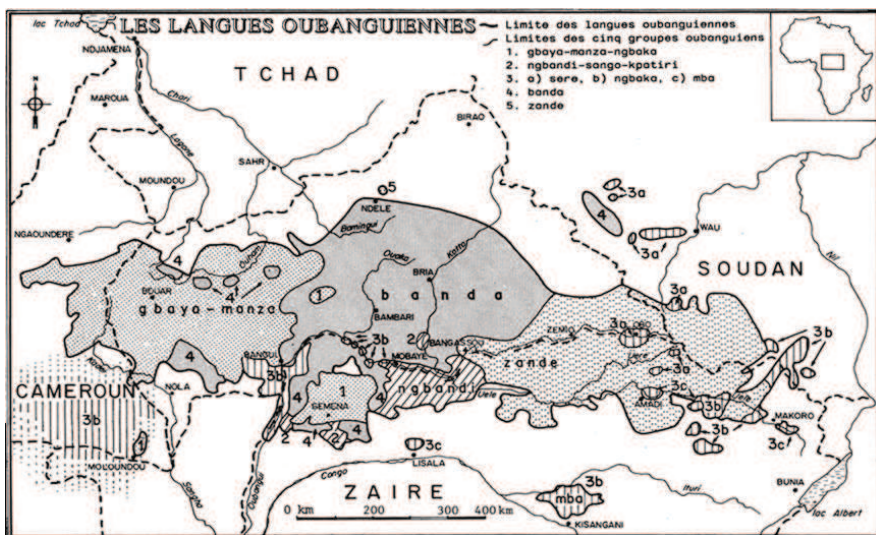
of high prominence within the CFA. In fact, I believe that the prominence of the region in southern Chad and the Central African Republic is epiphenomenal and that this region is not the core of the CFA. A much better candidate for this role is the Benue River corridor going from central Nigeria through northern Cameroon into southern Chad. The apparent prominence of this region within the CFA stems from a combined effect of a number of factors, primarily relevant to its southern part in the Central African Republic. On the one hand, we have the geography of the region, which makes it a kind of cul-de-sac with lots of marshy and seasonally flooded areas. On the other hand, we have the linguistic and population history of the region, which appears to be characterized by strong founder effects.

When looking closer into this part of the CFA, it is first important to note that this region of Central Africa is very homogenous with respect to the feature CFNM with most (if not all) languages having canonical CFNMs (type 4). At the same time, this region of Central Africa is both rather homogenous linguistically and rather sparsely populated. It is sparsely populated both in absolute terms, as becomes obvious from the low density of populated places in this region in Figure 13, and in terms of the languages spoken there. The latter fact is apparent in Figure 5, which shows the distribution of the languages of the sample.

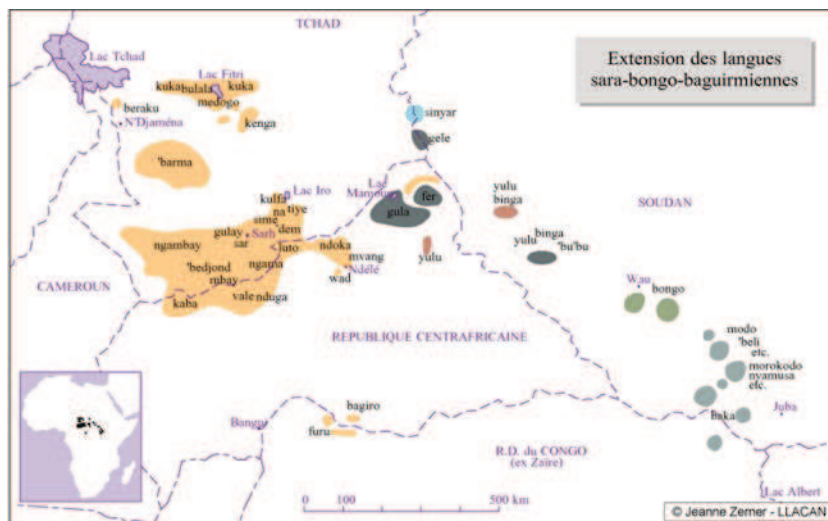


**Fig. 13:** Populated places in NSSA (based on the data from GeoNames.org) – the oval centered on the Central African Republic roughly indicates the relevant southern part of the CFA

To see how the sparse language density can be relevant for the apparent prominence of the region in question, recall the discussion of some of the possible artefacts of the visualization by means of spatial interpolation in Section 5.1. As for the linguistic homogeneity, this region is occupied by a small number of linguistic groups, all of which are rather shallow, viz., Gbaya-Manza-Ngbaka, Sere-Ngbaka-Mba, Banda, Ngbandi-Mongoba-Kazibati, Zande and Western Sara-Bongo-Bagirmi. All but the last group have traditionally been classified together under the label Ubangian, although more recently Gbaya-Manza-Ngbaka was excluded from this group. Western Sara-Bongo-Bagirmi is a branch of the Sara-Bongo-Bagirmi family, which itself is a subgroup within Central Sudanic. The map in Figure 14 illustrates the location of the first five groups under their older grouping as Ubangian and the map in Figure 15 shows the distribution of the Sara-Bongo-Bagirmi languages as a whole.



**Fig. 14:** Ubangian groups (Sere-Ngbaka-Mba, Banda, Ngbandi-Mongoba-Kazibati, Zande) and Gbaya-Manza-Ngbaka (formerly also classified as Ubangian) (Moñino 1988)

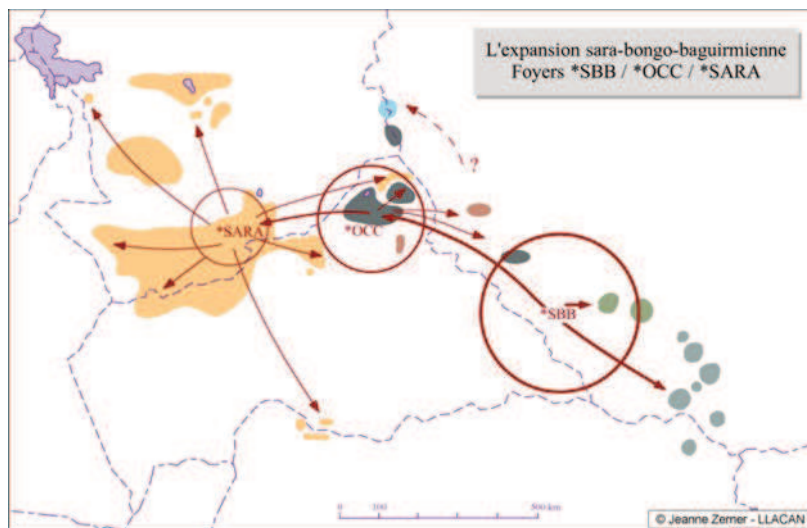


**Fig. 15:** Sara-Bongo-Bagirmi languages (Boyeldieu 2006) – the Western Sara-Bongo-Bagirmi languages are in pink (or light grey) (e.g., Kaba), dark grey (e.g., Gula or Gele) and brown (or normal grey) (e.g., Yulu)

The degree of internal diversity within all the six groups (Gbayya-Manza-Ngbaka, four Ubangian groups and Western Sara-Bongo-Bagirmi) is also rather low. Some of these groups could have just as well been referred to as languages without much exaggeration. Furthermore, within Ubangian, at least based on lexical similarity (cf. Boyeldieu and Cloarec-Heiss 1986; Moñino 1988:19), Sere-Ngbaka-Mba and Banda can be said to be rather closely related and probably also form one group with Ngbandi-Mongoba-Kazibati, with only Zande left out as not being transparently related to the other three Ubangian groups.

Another important point is that, beside being rather shallow and having a low level of internal diversity, most, if not all, of these groups are very likely to have moved into this region of Central Africa relatively recently and it is only upon their entering this region that most speciation events within these groups have occurred. For instance, Figure 16 shows the reconstructed migration routes of the Sara-Bongo-Bagirmi populations, indicating the Proto Sara-Bongo-Bagirmi homeland in what is now South Sudan just outside of our region in question, the Proto Western Sara-Bongo-Bagirmi diversification node in the northeast of the Central African Republic inside the region in question and an important node of further diversification within Western Sara-Bongo-Bagirmi around the border between Chad and the Central African Republic, viz., the Proto Sara node. Similarly,

as can be seen in Figure 14, both the Sere group of Sere-Ngbaka-Mba and a part of Banda are still spoken in South Sudan approximately in the same area as the Proto Sara-Bongo-Bagirmi homeland in Figure 16. Furthermore, the available evidence suggests that both the remaining part of Banda and the Ngbaka-Mba group of Sere-Ngbaka-Mba have also moved into the region in question from approximately the same area in South Sudan (e.g., Rombi and Thomas 2006: 22 for Ngbaka-Mba; Tisserant 1930: 8–10 for Banda).



**Fig. 16:** The reconstructed migration routes of the Sara-Bongo-Bagirmi populations (Boyeldieu 2006) (\*SBB is the Proto Sara-Bongo-Bagirmi homeland; \*OCC the Proto Western Sara-Bongo-Bagirmi diversification node; \*SARA the Proto Sara diversification node, a major subgroup within Western Sara-Bongo-Bagirmi)

Summing up, the region in question appears to have been occupied relatively recently by six shallow and rather homogenous linguistic groups and, of these six groups, at least three are transparently related to one another and may also be more distantly related to yet another group out of these six, which reduces the whole number of the different linguistic groups involved to three. These groups have moved into the region from outside. Moreover, all but one group have most likely migrated out of the same general area in what is now South Sudan, which makes it likely that they had been in close contact with each other even before entering the region in question. The only probable exception is Gbaya-Manza-

Ngbaka, which is more likely to have entered the region from the north somewhere in southern Chad and closer to the majority of the remaining languages of the CFA. These groups have then undergone further diversification upon entering the region. As the proto-language of Gbaya-Manza-Ngbaka is likely to have already had CFNMs, we end with a very high probability that the remaining two (or maximally three) proto-languages coming out of the same area simply happened to have CFNMs from the start or happened to acquire them upon entering the region. In such a situation, it is easy to imagine how the respective region may have easily become as homogenous as it is with respect to the relevant feature due to founder effects. In fact, the languages spoken in the region in question are rather homogenous with respect to a number of features that are otherwise highly unusual typologically,<sup>15</sup> such as the high lexical frequency of labial-velar stops (Idiatov and Van de Velde 2016, 2018), the prominent presence of labial flaps (Olson and Hajek 2003) and the use of possessee-like qualifier constructions (also known as dependency reversal) (Van de Velde 2012, 2013: 233–234).

In view of the arguments presented above, it is extremely unlikely that the prominent presence of canonical CFNMs in this region of Central Africa attests in any way to the hypothesized historical role of this region as a would-be primary core of the CFA. At the same time, as will be discussed in Section 5.3, this region of Central Africa with its prominent presence of CFNMs must have served as the source for the spread of the feature CFNM among Bantu languages further south in the Congo River corridor and the north of the Democratic Republic of Congo. Given the overall orientation of the CFA and its population dynamics as driven by the ecology and geography of the area, it is most likely that the primary historical core of the CFA is situated immediately to the northwest of the Central African Republic along the Benue River corridor going from southern Chad through northern Cameroon into central Nigeria. This is basically the high prominence region on the spatial intensity plot of languages with CFNMs in Africa in Figure 6. The region along the Benue River corridor is densely populated both in terms of people (cf. Figure 13) and languages (cf. Figures 5 and 6). Moreover, the linguistic landscape of this region is highly fragmented and characterized by a lot of deep linguistic diversity, which forms a stark contrast with the Central African Republic areas of the CFA further to the southeast that I have discussed first.

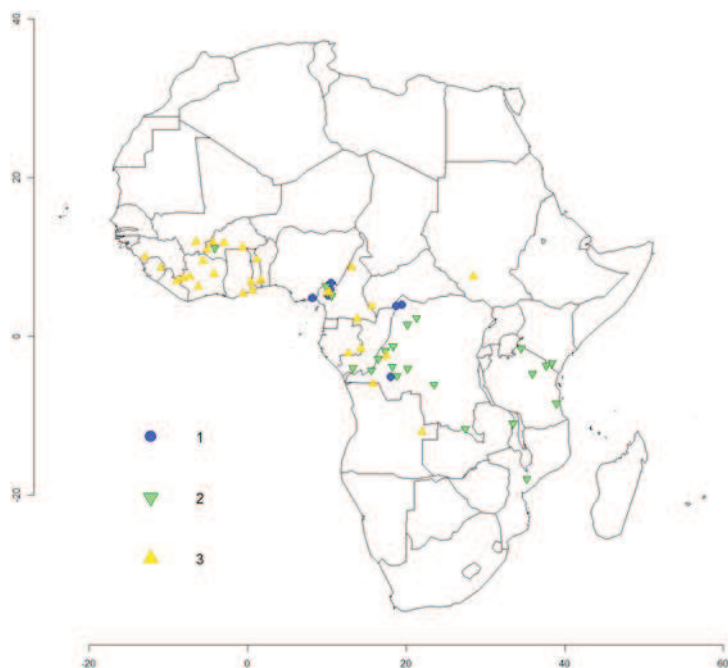
---

<sup>15</sup> This does not mean, of course, that these features need to be restricted to the languages of the region.

### 5.3 Optional and/or restricted CFNMs: grammatically non-canonical and areally peripheral

From the perspective of language change dynamics, optionality and constructional restrictions of CFNMs are likely to characterize either innovated or disappearing markers, depending on what the direction of change is. Detailed comparative studies would be required to determine the direction with certainty. Yet, given what I know about the languages in question, my strong impression is that such CFNMs are more often innovations than retentions from older stages on their way out.

From an areal typological perspective, languages with optional and constructionally restricted CFNMs, viz., languages with scores 1, 2 and 3 in the column “obligatoriness highest” in Table 1 (which I will refer to as types 1, 2, 3), are expected to be located mostly toward the periphery of the area of languages with CFNMs. This expectation is indeed borne out, as can be seen by comparing the spatial distribution of such languages in Figure 17 with that of all the languages having CFNMs in the sample in Figure 6 (cf. Section 3). Remarkably, languages of types 2 and 3 happen to be in almost complementary spatial distribution, with noticeable overlaps only in the Congo River corridor (cf. Section 3) and in southwestern Cameroon, which are also the same two regions where one finds languages of type 1. Moreover, languages of type 2 are almost all Bantoid and, within that group, they are almost all Narrow Bantu languages whereas the languages of type 3 are much more genetically diverse, which suggests that there is something different about Bantoid and especially Narrow Bantu languages. Typologically, Narrow Bantu languages are indeed known to differ in many respects from the languages in more northern parts of Sub-Saharan Africa, to most of which they are actually related (cf. Clements and Rialland 2008; Güldemann 2008).



**Fig. 17:** Languages with optional and constructionally restricted CFNMs (see Table 1 for the meanings of the values 1, 2, 3 in the column “obligatoriness highest”)

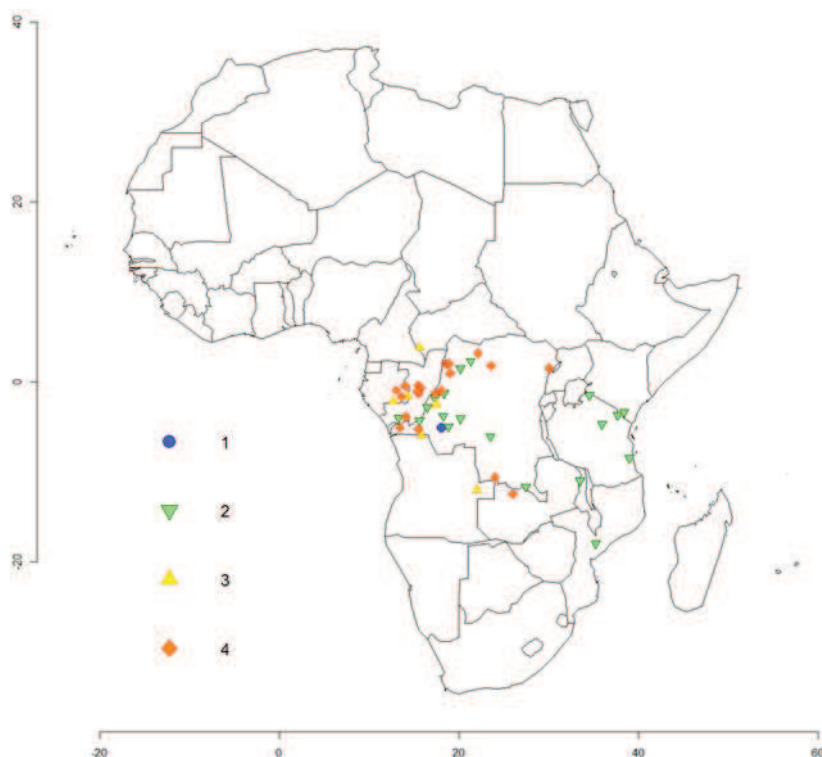
A possible explanation for the observed distribution of type 2 is that CFNMs in Bantoid and especially Narrow Bantu languages tend to develop through somewhat different pathways than elsewhere, the pathways that are less likely to lead to constructional restrictions but are more likely to result in optional CFNMs, either in the sense that they are clause-final but optional in that position or that they are optionally clause-final. As described by Devos and van der Auwera (2013), “recurrent sources for post-verbal negative markers [including CFNMs] in Bantu languages are locative pronouns, possessive pronouns and negative (answer) particles”, which indeed seem to be rarely attested as sources of CFNMs in more northern parts of Sub-Saharan Africa. As mentioned with respect to the negation marker *bo* in the Bantu language Nzadi in Section 2.5, which is optionally clause-final, possessive pronouns as a source of post-verbal negation markers are, for instance, unlikely to originate in the clause-final position. To some extent, the same is true for locative pronouns in Bantu. Negative answer particles as a source of post-verbal negation markers, although likely to originate in the clause-final position, are unlikely to be constructionally restricted.

The frequent optionality of CFNMs in Bantu must have much to do with their relatively young age. The relatively recent innovative character of CFNMs in Bantu is confirmed by their restricted distribution within Bantu and important variation in their forms across Bantu, which starkly contrasts with the relative uniformity and almost universal obligatory presence of the older pre-verbal negation markers (cf. Kamba Muzenga 1981; Güldemann 1999; Devos and van der Auwera 2013). Moreover, while the forms of the older pre-verbal negation markers can be reconstructed to Proto Bantu, they cannot be provided with an etymology other than a negation marker (Kamba Muzenga 1981). At the same time, CFNMs cannot be reconstructed to Proto Bantu and, when their etymology can be established, they often originate in elements that are not negation markers. The only noticeable exception are CFNMs originating in negative answer particles.

Another important factor contributing to the frequent optionality of CFNMs in Bantu is that, typically, clause-final markers as such are not a prominent morphosyntactic feature of Bantu languages and, in this respect, they clearly differ from languages of northern Sub-Saharan Africa (cf. Idiatov 2012a). Whereas the prominent presence of clause-final markers in the morphosyntax of languages of northern sub-Saharan Africa would be propitious for the upgrade of innovated CFNMs from optional to obligatory status, such a pull factor is generally lacking in Bantu languages.

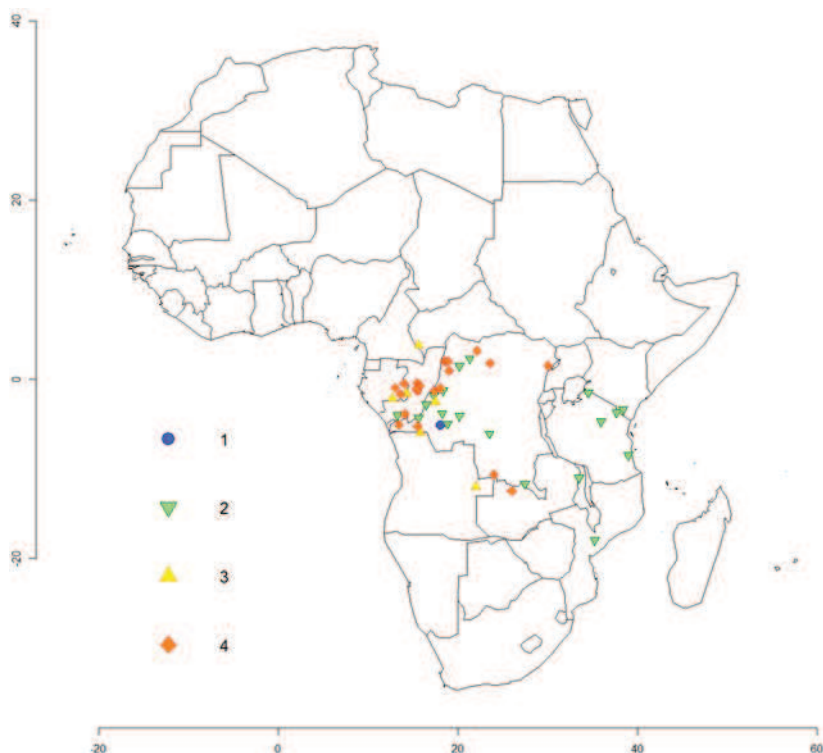
Within Bantu, the Congo River corridor is clearly the focal area of the innovation of CFNMs. This is suggested already by the observation in Figure 17 that Bantu languages of type 3 are basically confined to this region and that the concentration of Bantu languages of type 2 is also highest in the same region. The importance of the Congo River corridor becomes even more obvious when we include Bantu languages of type 4, i.e., Bantu languages with obligatory and constructionally unrestricted CFNMs. Thus, as can be seen in Figure 18, Bantu languages with more canonical CFNMs are equally concentrated in the Congo River corridor.





**Fig. 18:** Bantu languages with CFNMs (see Table 1 for the meanings of the values 1, 2, 3, 4 in the column “obligatoriness highest”)

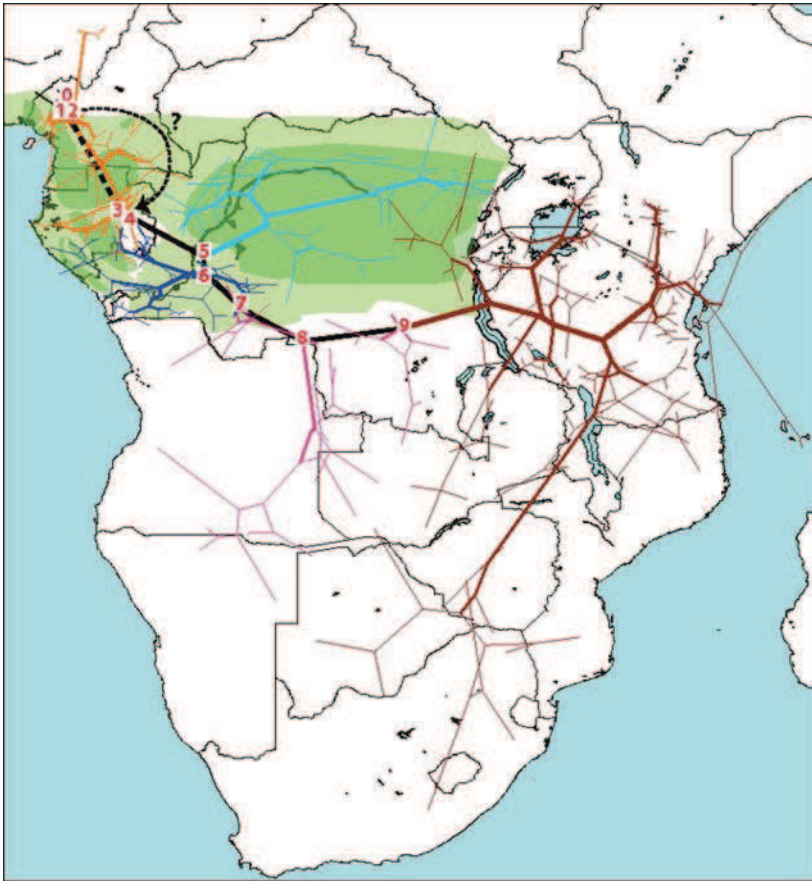
Furthermore, we can observe in Figure 18 two weak stretches of Bantu languages with CFNMs that appear to be linked to the northern and southern ends of the Congo River corridor and both going southeast, one to the north and the other to the south of the Congo River basin. Of these two secondary prominence zones, the southern one is clearly historically an offshoot of the Congo River corridor whereas the northern one must share its origin with the Congo River corridor in the CFA, situated further north in the Central African Republic (cf. Figures 8 and 11 and Section 5.2), as schematically illustrated in Figure 19. Admittedly, we cannot completely rule out the possibility that optional CFNMs in Bantu languages in East Africa have evolved independently.



**Fig. 19:** The suggested direction of spread of the use of CFNMs in Bantu from a major focal area of CFNM use in northern Central Africa into the Congo River corridor and the two secondary prominence zones

The emergence of both the Congo River corridor and the two secondary prominence zones must result from relatively recent population and/or language movements out of Central Africa. They clearly occurred much later than the original Bantu expansion in and around the Congo River basin. In this respect, compare the areal pattern of the distribution of CFNMs in Bantu with the Bantu expansion route reconstructed by Grollemund et al. (2015) and reproduced in Figure 20. The comparison makes it clear that the southwest expansion of the use of CFNMs in the Congo River corridor went in the direction opposite to the original route of Bantu expansion in the northern half of the Congo River basin. The northern secondary prominence zone does not correspond to any original route of Bantu expansion in that area from the northern Democratic Republic of Congo. The southern secondary prominence zone partially corresponds to an original route of Bantu expansion in that area, yet it could not have formed before the emergence

of the Congo River corridor zone and therefore could not have coincided in time with this part of the Bantu expansion route.



**Fig. 20:** Bantu migration route reconstructed by Grollemund et al. (2015) on consensus tree by using geographical locations of contemporary languages and connecting ancestral locations by straight lines (the true route will differ) (numbered positions correspond to major diversification nodes on the consensus tree; the curved dashed line indicates the suggested migration route through savannah corridors; lighter (green) shading corresponds to the delimitation of the rainforest at 5,000 BP.; the darker (green) shading corresponds to the delimitation of the rainforest at 2,500 BP)

## 6 Concluding remarks

The synchronic patterns that we discover are necessarily a product of language change as it evolves in time and space. This paper provides an analysis of spatio-temporal language dynamics in Sub-Saharan Africa with respect to the feature CFNM. It is my strong conviction that the most plausible account of synchronic patterns can only be gleaned by casting your net wide to catch more of the synchronic diversity, rather than by trying to reduce it. Furthermore, when analyzing areal patterns, it is also important to consider together the languages that have the feature under investigation and the languages that do not have it. In terms of spatial analysis, methods such as spatial interpolation and generalized additive modeling (including their mixed extensions) provide particularly valuable research tools.

A question that one is often tempted to ask when doing areal typology is which linguistic group of the ones present in the area where the feature is prominent could have been the primary vector of the feature. Yet, this question can only have a fully meaningful answer if we know that no linguistic groups have disappeared from the scene without traces since the emergence of the feature in the area. Unfortunately, in a region such as Sub-Saharan Africa, we cannot be sure of that. In fact, we can be quite sure of the contrary.<sup>16</sup> Furthermore, the fact that all members of a certain linguistic group carry the feature and are spoken inside the area where the feature is prominently present cannot prove that this linguistic group is the primary vector of the feature.

For instance, Dryer (2009: 346) entertains two possible scenarios with respect to his core area of VO&VNeg languages in Central Africa, which, unsurprisingly, largely coincides with our CFA of the feature CFNM.<sup>17</sup> The first scenario is that the feature VO&VNeg originates in Chadic languages, because the feature is pervasive in Chadic languages and all Chadic languages are spoken inside the relevant

---

<sup>16</sup> For instance, see Kleinewillinghöfer (2001) on Jalaa, an apparent linguistic isolate in north-eastern Nigeria, the area that is particularly relevant for the feature CFNM, which went extinct fairly recently and for which only some lexical data could be collected from rememberer speakers.

<sup>17</sup> It is not very surprising because Dryer (2009) restricts his study to negation markers that are words. Although not all post-verbal negation markers that are words are also clause-final, CFNMs are almost always analyzed as words precisely because of their clausal orientation and, therefore, at least canonical CFNMs would always be classified as post-verbal negation markers in Dryer's (2009) typology.

area. Yet, Chadic languages are also typologically quite different from their distant Afro-Asiatic relatives, most of which equally do not carry the relevant feature. In other words, we cannot know whether the feature can be reconstructed to Proto Chadic as independent of the current participation of Chadic languages in the area in question. The second scenario is that the presence of the feature in Chadic is due to a substrate influence from “Nilo-Saharan” and that the substrate influence could have such a pervasive effect on Chadic languages because of the relatively small size of the region populated by speakers of Chadic languages. The Nilo-Saharan group that is both spoken in the same area as Chadic languages and is classified positive for the relevant feature by Dryer (2009: 311) is Western Sara-Bongo-Bagirmi. However, as laid out in Section 5.2, Western Sara-Bongo-Bagirmi languages are one of the recent newcomers in the area. That is, even if there was a substrate influence on Chadic, which is actually quite plausible given the typological differences between Chadic and its Afro-Asiatic relatives, we cannot know what that substrate was, nor can we know whether the pervasive presence of the feature in Chadic can be attributed to this substrate.

The considerations above equally apply to the role of Chadic and other similarly homogenous groups (such as Gbaya-Manza-Ngbaka and Zande) within our CFA of the feature CFNM. While a homogenous distribution of the feature within a given linguistic group (all members of the group carry the feature and are spoken inside the area) does not tell us much on the spatio-temporal dynamics of the feature, a much more informative signal is usually provided by groups that are diverse with respect to the feature, with members both inside and outside the area, especially when it can be complemented by independent information on language and population movements. Thus, in the case of the feature CFNM, various Niger-Congo groups are spoken around the Benue River corridor in the CFA, as well as further to the west in the WFA. At the same time, many Niger-Congo groups are also spoken outside of the CFA and the WFA. We can reasonably hypothesize that the latter Niger-Congo groups have lost the feature CFNM when they moved outside of the area, as when entering the forest zone along the coast of the Gulf of Guinea, for instance, due to some substrate influence (just as, for instance, they developed high lexical frequency of labial-velars in the same coastal regions; cf. Idiatov and Van de Velde 2016, 2018), or when some groups lacking the feature entered the area from outside. For instance, as discussed in Section 5.1, the former type of loss is likely to be the case for the Yoruboid languages in the coastal gap in southwestern Nigeria between two coastal extensions of the CFA while the latter type of loss is likely to be the reason behind the emergence of the major discontinuity separating the WFA from the CFA. At the same time, it is rather unlikely that the feature CFNM should be reconstructed to

higher nodes in the Niger-Congo tree, not even to the Proto Benue-Congo node. A major counterargument to such a deep reconstruction is presented by the general lack of CFNMs in Southern Bantoid languages (with the most noticeable exception of the Bantu languages of the Congo River corridor but, as discussed in Section 5.3, this is clearly a recent development). Within Benue-Congo, their superordinate group, and within Niger-Congo in general, Southern Bantoid languages and especially Bantu languages are generally considered archaic in their typological profile (e.g., Hyman 2011).

**Acknowledgement:** This work is situated within the project LC2 *Areal phenomena in Northern Sub-Saharan Africa* of the Labex EFL (the *Investissements d'Avenir* program, overseen by the French National Research Agency; reference: ANR-10-LABX-0083). Special thanks with respect to the present paper are due to Mark Van de Velde. I am grateful to the referees and the editors for their constructive criticism. Last but not least, my sincere gratitude goes to Johan van der Auwera, who first sparked my interest in the study of negation. Johan got me started in linguistic typology and has always been a great example to me in many respects.

## References

- Andersen, Torben. 1981. *A grammar of Modo: A preliminary sketch*. Aalborg: University Center of Aalborg.
- Baayen, R. Harald. 2013. Multivariate statistics. In Robert J. Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 337–372. Cambridge: Cambridge University Press.
- Baddeley, Adrian & Rolf Turner. 2005. spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software* 12 (6). 1–42.
- Beyer, Klaus. 2009. Double negation marking: A case of contact-induced grammaticalization in West-Africa? In Norbert Cyffer, Erwin Ebermann & Georg Ziegelmeyer (eds.), *Negation patterns in West African languages and beyond*, 205–222. Amsterdam: John Benjamins.
- Bisang, Walter & Remi Sonaiya. 2000. Information structuring in Yoruba. *Linguistics* 38 (1). 169–197.
- Boyeldieu, Pascal. 2006. *Présentation des langues Sara-Bongo-Baguirmiennes*. Paris: CNRS-LLACAN. <http://sumale.vjf.cnrs.fr/SBB/presentation.html> (accessed 7 April 2018).
- Boyeldieu, Pascal & France Cloarec-Heiss. 1986. Dialectometrie lexicale dans le domaine ou-banguien. In Gladys Guarisma & Wilhelm J. G. Möhlig (eds.), *La méthode dialectométrique appliquée aux langues africaines*, 331–393. Berlin: Dietrich Reimer Verlag.
- Brown, Dunstan, Marina Chumakina & Greville G. Corbett (eds.). 2013. *Canonical morphology and syntax*. Oxford: Oxford University Press.
- Clements, G.N. & Annie Rialland. 2008. Africa as a phonological area. In Bernd Heine & Derek Nurse (eds.), *A linguistic geography of Africa*, 36–85. Cambridge: Cambridge University Press.

- Crane, Thera M., Larry M. Hyman & Simon Nsielanga Tukumu. 2011. *A grammar of Nzadi [B865]: A Bantu language of Democratic Republic of Congo*. Berkeley: University of California Press.
- Creissels, Denis. 2005. S-O-V-X constituent order and constituent order alternations in West African languages. In Rebecca Cover & Yuni Kim (eds.), *Proceedings of the Berkeley Linguistics Society 31st annual meeting*, 37–51. Berkeley: University of California at Berkeley.
- Croft, William. 1991. The evolution of negation. *Journal of Linguistics* 27 (1). 1–27.
- Cyffer, Norbert, Erwin Ebermann & Georg Ziegelmeyer. 2009. *Negation patterns in West African languages and beyond*. Amsterdam: John Benjamins.
- Devos, Maud & Johan van der Auwera. 2013. Jespersen Cycles in Bantu: Double and triple negation. *Journal of African Languages and Linguistics* 34 (2). 205–274.
- Dryer, Matthew S. 2009. Negation patterns in West African languages and beyond. In Norbert Cyffer, Erwin Ebermann & Georg Ziegelmeyer (eds.), *Verb-object-negative order in Central Africa*, 307–362. Amsterdam: John Benjamins.
- Grollemund, Rebecca, Simon Branford, Koen Bostoen, Andrew Meade, Chris Venditti & Mark Pagel. 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences* 112 (43). 13296–13301.
- Güldemann, Tom. 1999. The genesis of verbal negation in Bantu and its dependency on functional features and clause types. In Jean-Marie Hombert & Larry M. Hyman (eds.), *Bantu historical linguistics: Theoretical and empirical linguistics*, 545–587. Stanford: Center for the Study of Language and Information.
- Güldemann, Tom. 2008. The Macro-Sudan belt: Towards identifying a linguistic area in northern sub-Saharan Africa. In Bernd Heine & Derek Nurse (eds.), *A linguistic geography of Africa*, 151–185. Cambridge: Cambridge University Press.
- Heath, Jeffrey. 2008. *A grammar of Jamsay*. Berlin: De Gruyter.
- Hyman, Larry M. 2011. The Macro-Sudan belt and Niger-Congo reconstruction. *Language Dynamics and Change* 1. 3–49.
- Idiatov, Dmitry. 2008. Antigrammaticalization, antimorphologization and the case of Tura. In Elena Seoane, María José López-Couso & Teresa Fanego (eds.), *Theoretical and empirical issues in grammaticalization*, 151–169. Amsterdam: John Benjamins.
- Idiatov, Dmitry. 2012a. On the history of clause-final negation in the Mande languages of the Bani – upper Mouhoun rivers area. Paper presented at the workshop “The history of post-verbal negation in African languages” at the 7th World Congress of African Linguistics, University of Buea, 20–24 August. [http://idiatov.mardi.myds.me/WOCAL7\\_Negation/IDI-ATOV\\_2012\\_Presentation.pdf](http://idiatov.mardi.myds.me/WOCAL7_Negation/IDI-ATOV_2012_Presentation.pdf) (accessed 7 April 2018).
- Idiatov, Dmitry. 2012b. Clause-final negative markers in southeastern Bamana dialects: A contact-induced evolution. *Africana Linguistica* 18. 169–192.
- Idiatov, Dmitry. 2015. Clause-final negative markers in Bobo and Samogo: Parallel evolution and contact. *Journal of Historical Linguistics* 5 (2). 235–266.
- Idiatov, Dmitry. in prep. Clause-final negation in northern sub-Saharan Africa: Right periphery, intersubjectivity and areality.
- Idiatov, Dmitry & Mark Van de Velde. 2016. The lexical frequency of labial-velar stops in northern sub-Saharan Africa and its historical implications. Paper presented at the workshop “Areal features and linguistic reconstruction in Africa” at the 47th Annual Conference on African Linguistics, University of California at Berkeley, 23–26 March. <https://halshs.archives-ouvertes.fr/halshs-01481552> (accessed 7 April 2018).

- Idiatov, Dmitry & Mark Van de Velde. 2018. La fréquence lexicale des occlusives labiales-vélaires dans le nord de l'Afrique sub-saharienne. In Jean-Léo Léonard & Annie Rialland (eds.), *Linguistique africaine: Perspectives croisées*. Leuven: Peeters.
- Ingham, Richard. 2013. Negation in the history of English. In David Willis, Christopher Lucas & Anne Breitbarth (eds.), *The history of negation in the languages of Europe and the Mediterranean*, 119–148. Oxford: Oxford University Press.
- Jones, Ross McCallum. 1998. *The Boko/Busa language cluster*. Munich: LINCOM Europa.
- Kamba Muzenga, J. G. 1981. *Les formes verbales négatives dans les langues bantoues*. Ter-vuren: Royal Museum for Central Africa.
- Kleinewillinghöfer, Ulrich. 2001. Jalaa, an almost forgotten language of northeastern Nigeria: A language isolate? *Sprache und Geschichte in Afrika* 16/17. 239–271.
- Littig, Sabine & Ulrich Kleinewillinghöfer. 2012. Negation patterns in Sama-Duru languages (Central Adamawa). Paper presented at the workshop “The history of post-verbal negation in African languages” at the 7th World Congress of African Linguistics, University of Buea, 20–24 August. [http://idiatov.mardi.myds.me/WOCAL7\\_Negation/LITTIG\\_KLEIN-EWILLINGHOEFER\\_2012\\_Presentation.pdf](http://idiatov.mardi.myds.me/WOCAL7_Negation/LITTIG_KLEIN-EWILLINGHOEFER_2012_Presentation.pdf) (accessed 7 April 2018).
- Lucas, Christopher. 2009. *The development of negation in Arabic and Afro-Asiatic*. Cambridge: University of Cambridge PhD dissertation.
- Maniacky, Jacky. 2003. *Tonologie du ngangela: Variété de Menongue (Angola)*. Munich: LINCOM.
- Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.
- Miestamo, Matti. 2008. *Standard negation: The negation of declarative verbal main clauses in a typological perspective*. Berlin: De Gruyter.
- Moñino, Yves. 1988. Introduction: Cousines ou voisines? In Yves Moñino (ed.), *Lexique comparatif des langues oubanguiennes*, 11–22. Paris: Geuthner.
- Olson, Kenneth S. & John Hajek. 2003. Crosslinguistic insights on the labial flap. *Linguistic Typology* 7 (2). 157–186.
- Persson, Andrew M. & Janet R. Persson. 1991. *Möddö-English dictionary with grammar*. Nairobi: Summer Institute of Linguistics.
- R Core Team. 2015. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org> (accessed 7 April 2018).
- Rombi, Marie-Françoise & Jacqueline M. C. Thomas. 2006. *Un continuum prédicatif: Le cas du Gbanzili (République Centrafricaine)*. Paris: Peeters.
- Roulon-Doko, Paulette. 2012. Le marqueur de négation ná en gbaya. Paper presented at the workshop “The history of post-verbal negation in African languages” at the 7th World Congress of African Linguistics, University of Buea, 20–24 August. [http://idiatov.mardi.myds.me/WOCAL7\\_Negation/ROULON-DOKO\\_2012\\_Presentation.pdf](http://idiatov.mardi.myds.me/WOCAL7_Negation/ROULON-DOKO_2012_Presentation.pdf) (accessed 7 April 2018).
- Solomiac, Paul. 2007. *Phonologie et morphosyntaxe du dzùùngoo de Samogohiri*. Lyon: Université Lumière Lyon 2 PhD dissertation.
- Tamminga, Meredith, Christopher Ahern & Aaron Ecay. 2016. Generalized Additive Mixed Models for intraspeaker variation. *Linguistics Vanguard* 2 (s1). 1–9.
- Tisserant, Charles. 1930. *Essai sur la grammaire banda*. Paris: Institut d'Ethnologie.
- Tosco, Mauro. 2001. *The Dhaasanac language: Grammar, texts, vocabulary of a Cushitic language of Ethiopia*. Cologne: Rüdiger Köppe.
- Van de Velde, Mark. 2012. The origin and spread of possessee-like qualifiers in Central Africa. Paper presented at the 7th World Congress of African Languages, University of Buea, 20–



- 24 August. [http://llacan.vjf.cnrs.fr/pers/vandevelde/files/Van\\_de\\_Velde\\_WOCAL\\_BUEA.pptx](http://llacan.vjf.cnrs.fr/pers/vandevelde/files/Van_de_Velde_WOCAL_BUEA.pptx) (accessed 7 April 2018).
- Van de Velde, Mark. 2013. The Bantu connective construction. In Anne Carlier & Jean-Christophe Verstraete (eds.), *The genitive*, 217–252. Amsterdam: John Benjamins.
- van der Auwera, Johan. 2009. The Jespersen cycles. In Elly van Gelderen (ed.), *Cyclical change*, 35–71. Amsterdam: John Benjamins.
- van der Auwera, Johan. 2010. On the diachrony of negation. In Laurence R. Horn (ed.), *The expression of negation*, 73–101. Berlin: De Gruyter.
- van der Auwera, Johan. & Lauren Van Alsenoy. 2016. On the typology of negative concord. *Studies in Language* 40 (3). 473–512.
- Van Gelderen, Elly. 2008. Negative cycles. *Linguistic Typology* 12 (2). 195–243.
- Veselinova, Ljuba. 2013. Negative existentials: A cross-linguistic study. *Rivista di Linguistica* 25 (1). 107–145.
- Wieling, Martijn, Simonetta Montemagni, John Nerbonne & R. Harald Baayen. 2014. Lexical differences between Tuscan dialects and Standard Italian: Accounting for geographic and sociodemographic variation using Generalized Additive Mixed Modeling. *Language* 90 (3). 669–692.
- Wieling, Martijn, John Nerbonne & R. Harald Baayen. 2011. Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE* 6 (9). e23613.
- Winkelmann, Kerstin & Gudrun Mieke. 2009. Negation in Gur: Genetic, areal and unique features. In Norbert Cyffer, Erwin Ebermann & Georg Ziegelmeyer (eds.), *Negation patterns in West African languages and beyond*, 167–204. Amsterdam: John Benjamins.
- Winter, Bodo & Martijn Wieling. 2016. How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution* 1 (1). 7–18.
- Wood, Simon N. 2015. *mgcv*. <http://CRAN.R-project.org/package=mgcv> (accessed 7 April 2018).