

## **The Distribution of Advanced Tongue Root Harmony and Interior Vowels in the Macro-Sudan Belt**

Nicholas Rolle, Matthew Faytak & Florian Lionnet\*

**Abstract.** In this paper we investigate the distribution of vowel systems in the Macro-Sudan Belt, an area of Western and Central Africa proposed in recent areal work (Güldemann 2008, 2011; Clements & Rialland 2008). We report on a survey of 615 language varieties with entries coded for two phonological features: advanced tongue root (ATR) harmony and the presence of interior vowels (i.e. non-peripheral vowels, such as [i u ɜ ə ʌ ...]). Our results show that the presence of ATR harmony in the Macro-Sudan Belt is limited to three separated zones: an Atlantic ATR Zone, a West African ATR Zone, and an East African ATR Zone, all geographically unconnected to one another. We additionally show that between the West and East African ATR Zones is a geographically extensive, genetically heterogeneous region of Central Africa where ATR harmony is systematically absent which we term the Central African ATR-less Zone. Our results also show a large region where phonemic and allophonic interior vowels are disproportionately prevalent, which we term the Central African Interior Vowel Zone. This zone noticeably overlaps with the Central African ATR-less Zone, suggesting that ATR and interiority have an antagonistic relationship. Chi-squared tests support the presence of a strong relationship between the two types of vowel contrasts.

**Keywords.** areal linguistics; advanced tongue root harmony; interior vowels; phonological database; typology; West Africa; Central Africa

**1. Introduction.** This paper explores the distribution of vowel system properties within the Macro-Sudan Belt, a proposed macro-linguistic area of West and Central Africa identified in Güldemann (2008) and independently called the Sudanic Belt in Clements & Rialland (2008). This proposed macro-area encompasses a vast geographic region stretching from Senegal to South Sudan south of the Sahara Desert and north of the Congo Rainforest. One of the features argued to define the Macro-Sudan Belt as a linguistic area is the presence of Advanced Tongue Root (ATR) harmony (Stewart 1967, Dimmendaal 2001, Casali 2003, 2008, Clements & Rialland 2008, Güldemann 2008). Casali (2003) in particular presents a thorough overview of the types of ATR harmony systems and the phonological properties they exhibit, and demonstrates that ATR harmony cuts across language families in Africa, strongly suggesting areal convergence. Of all the linguistic features thought to define the Macro-Sudan Belt, ATR

\* This paper stems from a working group at the University of California, Berkeley called the ALFA Project (Areal Linguistic Features of Africa). We would like to thank our colleagues on this project for valuable feedback: Larry Hyman, Jack Merrill, Emily Clem, Hannah Sande, Peter Jenks, and Nico Baier. Further thanks are due to Lev Michael, Dmitry Idiatov, Sharon Rose, Jeremy Steffman, and the audience of the 2017 LSA, including Bernard Comrie, Jeff Good, Bonny Sands, Ron Schaefer, and Krystyna Wachowicz. Previous versions of this paper were delivered at the 47<sup>th</sup> Annual Conference on African Linguistics (UC Berkeley, 2016) and at the Colloquium on African Languages and Linguistics (Universiteit Leiden, 2016). We would like to thank those audiences as well. Many linguists provided data for individual languages for which we are very grateful. Authors: Nicholas Rolle, UC Berkeley ([n.rolle@berkeley.edu](mailto:n.rolle@berkeley.edu)), Matthew Faytak, UC Berkeley ([mf@berkeley.edu](mailto:mf@berkeley.edu)) & Florian Lionnet, UC Berkeley/Princeton University ([flionnet@princeton.edu](mailto:flionnet@princeton.edu)).

harmony is the most widespread, found in Atlantic languages in the far west, Somali in the far east (Saeed 1999), and Malila in southern Tanzania (Kutsch Lojenga 2009).

At the same time, additional literature has shown that many languages within the Macro-Sudan Belt lack ATR contrasts and harmony, many of which appear in a large region of Central Africa (Boyd 1989:197, Dimmendaal 2001:370). In this Central region, other vowel system tendencies have been noted, such as the prevalence of highly restricted vowel co-occurrence phonotactics within morphemes and words that are not reducible to ATR harmony, e.g. in Mbay [myb] (Sara, Central-Sudanic: Chad – Keegan 1997). Further, one phonological feature which does coincide with this Central area is the widespread presence of interior vowels, i.e. non-peripheral vowels such as [i u ɜ ə ʌ ...], noted to our knowledge only by Thomas et al. (1973).

In this paper, we present findings from an areal-typological survey using a database of 615 language varieties within the Macro-Sudan Belt (as well as in adjacent areas), coding for ATR harmony and interiority. For ATR we code languages for three values: ‘Strict ATR’, ‘Trace ATR’ and ‘No ATR’. For interiority we code for four values: ‘Present – Phonemic’, ‘Present – Non-phonemic’, ‘Present – [+ATR, +low] V only’, and ‘Absent – No interior vowels’. Our results show ATR prevalent in three geographically unconnected zones: an Atlantic ATR Zone centered around Senegal, a West African ATR Zone along the West African shore of the Gulf of Guinea and further inland, and an East African ATR Zone stretching from Northern Chad to Tanzania. Between the West and East Zones, we point to the existence of an expansive region of Central Africa systematically lacking ATR contrasts or harmony, which we call the Central African ATR-less Zone. Furthermore, our results also show a large region where phonemic and allophonic interior vowels are disproportionately prevalent, which we term the Central African Interior Vowel Zone. These two linguistic areas overlap substantially, covering a genetically heterogeneous set of languages including Grassfields Bantoid, Bantu A, East Benue-Congo, Kainji, Platoid, Jukunoid, Adamawa, Ubangi, Central Sudanic, and Chadic. In contrast, we identify a West African Interior Vowel-less Zone extending from Guinea to central Nigeria which largely overlaps with the West African ATR Zone.

Our survey results reveal that the most extensively discussed areas of ATR in Africa are in fact separated by a vast ATR-less region, and that there are in fact two distinct vowel system profiles in the Macro-Sudan Belt, complicating its status as a linguistic area. The strong inverse relationship between the presence of ATR harmony and interiority is supported with a chi-square test that shows a significant relationship between the two variables. We interpret these results as showing that elaboration of phonological contrast along the acoustic dimension of F1 as in ATR harmony is antagonistic to elaboration of interiority distinctions along the F2 dimension. The structure of this paper is as follows. Section 2 presents our survey of 615 African language varieties found in the Macro-Sudan Belt, showing our results on the distribution of ATR and interiority. Section 3 provides discussion of the inverse relationship between ATR and interiority. Section 4 provides concluding remarks.

**2. Survey of vowel systems.** Our current survey is a selection of 615 language varieties from our larger database on African vowel systems. This vowel database seeks to attain complete coverage of all major language varieties in the Macro-Sudan Belt and surrounding environs, limited only by existence, access, and reliability of relevant phonological descriptions. Data collection for this vowel database was done by all three authors through a brute-force search of phonological descriptions in the literature, avoiding wordlists in favor of grammars and/or phonological sketches. For some languages, only wordlists were available, and we include them if we surmise that they are of sufficient quality to deduce the phonological structure of the

language. We restricted our research to primary sources and did not consult existing databases, i.e. PHOIBLE (Moran et al. 2014), the World Phonotactics Database (Donohue et al. 2013), *Alphabets des langues africaines* (Hartell 1993), UPSID (Maddieson 1984), etc.

For each language, ISO 639 codes, genetic affiliation, and geographic location were obtained from Glottolog (Hammarström et al. 2016), and the language-specific literature was used to determine each language’s complete vowel inventory, type of ATR harmony (if applicable), and the presence of other types of harmonies (height, front/back, rounding). Each language’s vowel inventory was coded for a variety of additional contrasts and secondary articulations (vowel length, nasality, glottalization, breathiness, and pharyngealization). Additionally, we included in our coding of vowel inventories a distinction between phonemic vowel contrasts and allophonic surface variants. For instance, the vowel inventory of Eton (Bantu A: Cameroon – van de Velde 2008) was coded with two surface variants [ɛ, ɜ] of the phoneme /ɛ/, the latter of which occurs in closed syllables. Finally, we also coded whether a given vowel is epenthetic, whether it is the reduced variant of any other vowels, and whether the vowel is marginal in the language (e.g. found in only a handful of morphemes).

Table 1 provides the number of languages from each major genetic family in and abutting the Macro-Sudan Belt which we surveyed. Figure 1 provides the locations of all language varieties used in the current survey (red circles, n = 615). The transparent circles represent varieties not included in the current survey, for which we do not have sufficient data.

| Family              | #  | Family      | #  | Family        | #  | Family        | # |
|---------------------|----|-------------|----|---------------|----|---------------|---|
| Bantu               | 70 | Nilotic     | 17 | Other Bantoid | 13 | North Bantoid | 8 |
| Central Sudanic     | 36 | Other NC    | 17 | West Mande    | 13 | West Kru      | 8 |
| Other NS            | 31 | West Chadic | 17 | Other AA      | 12 | Idomoid       | 7 |
| Atlantic            | 26 | Biu-Mandara | 15 | East Chadic   | 10 | Igboid        | 7 |
| Adamawa             | 25 | Platoid     | 15 | Ijoid         | 10 | Omotic        | 7 |
| Potou-Tano          | 25 | East Mande  | 14 | Kordofanian   | 10 | Defoid        | 6 |
| Ubangi              | 23 | Kainji      | 14 | Delta-Cross   | 9  | Senoufo       | 6 |
| Edoid               | 22 | Other Kwa   | 14 | Gbe           | 9  | Songhai       | 5 |
| Central Gur         | 21 | South Gur   | 14 | Jukunoid      | 9  | Dogon         | 4 |
| Grassfields Bantoid | 20 | Gbaya       | 13 | Surmic        | 9  | Other         | 4 |

Table 1: Languages surveyed from each family (n = 615)

Grey = Niger-Congo (NC), dark grey = controversially NC, blue = Nilo-Saharan (NS), orange = Afro-Asiatic (AA)

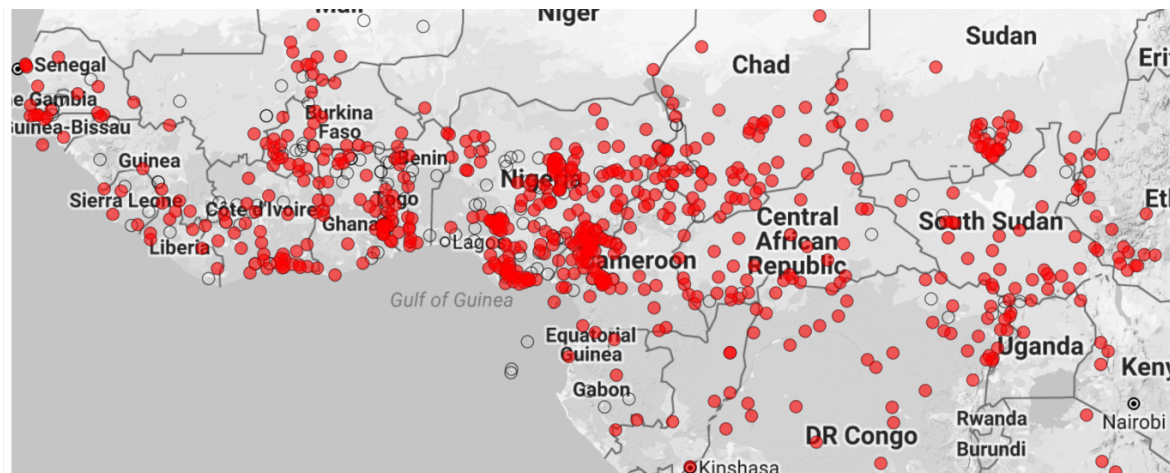


Figure 1: Language varieties in this survey (red circles, n = 615)

2.1. SURVEY OF ATR HARMONY. Other than tone, advanced tongue root (ATR) contrasts and ATR harmony are perhaps the most robustly analyzed features of African phonological systems. In such systems, vowels are split into two mutually exclusive groups. In the [+ATR] group, vowels show advancement of the tongue root which widens the pharyngeal cavity, whereas [-ATR] vowels (also called retracted tongue root [RTR] vowels) do not show advancement. Acoustically, [+ATR] vowels canonically have a lower first formant frequency (F1) than their [-ATR] counterparts. [+ATR] vowels are marked with a diacritic /̠/ and [-ATR] with /̡/. As F1 is also the primary cue for tongue height, there is also a convention of [+ATR] vowels being transcribed with higher/more closed segments compared to their [-ATR] counterparts, e.g. [e̠] vs. [e̡].<sup>1</sup> An example of a vowel system with a full set of ATR contrasts is provided below from Degema [deg] (Edoid: Nigeria – Kari 2007).

- (1) Degema ATR contrasts
- |                               |                               |
|-------------------------------|-------------------------------|
| [+ATR]                        | [-ATR]                        |
| /i̠ e̠ ɔ̠ u̠/ ~ /i̡ e̡ ɔ̡ u̡/ | /ɪ̠ ɛ̠ ɔ̠ ʊ̠/ ~ /ɪ̡ ɛ̡ ɔ̡ ʊ̡/ |

ATR systems display ATR harmony in which vowels co-occur only with other vowels in their set. Canonically, there are both vowel co-occurrence restrictions within roots (static patterns) and restrictions across morphemes within a word resulting in allomorphy (dynamic patterns). In (2) below, a. shows that [+ATR] and [-ATR] vowels do not co-occur within the same root, and b. shows that they also do not co-occur within a word resulting in allomorphy of the suffix ([-ATR] is indicated with a dot below the first vowel of the word in Degema orthography).

- (2) ATR harmony in Degema
- a. Static: Vowel co-occurrence constraints in roots
- |        |              |  |         |
|--------|--------------|--|---------|
| [+ATR] | vuné /β̀ùné/ | ‘slit stomach and take out intestines’ | (*vunɛ) |
| [-ATR] | mùré /m̀òré/ | ‘light (a fire)’                       | (*more) |
- b. Dynamic: Vowel co-occurrence constraints across morphemes showing allomorphy
- |        |               |         |                    |                        |
|--------|---------------|---------|--------------------|------------------------|
| [+ATR] | dúw ‘be soft’ | duw-esé | ‘cause to be soft’ | [d̀ùw-èsé] (*d̀ùw-èsé) |
| [-ATR] | sín ‘climb’   | sìn-esé | ‘cause to climb’   | [s̀ìn-èsé] (*s̀ìn-èsé) |

<sup>1</sup> Casali (2008:510) notes that [ATR] contrasts may additionally be cued through voice quality. [+ATR] are described as “breathy”, “deep”, “muffled”, or “hollow” quality, whereas [-ATR] are described as “bright”, “choked”, “brassy”, or “creaky”.

Numerous Africanists have discussed the areal distribution of ATR across the continent. Güldemann (2008) maps a near-continuous band of ATR stretching from Senegal to Ethiopia, the only gap being in the Mande area centered around Guinea and western Mali. Working independently, Clements & Rialland (2008) summarize the occurrence of ATR harmony and observe a notable absence of ATR among Atlantic families, Gbe, Defoid (i.e. Yoruboid), Idomoid, Platoid, Jukunoid, Bantoid/Northwestern Bantu languages (zones A–D), and Adamawa-Ubangi (except Zandic). Clements & Rialland’s more conservative estimate of ATR distribution aligns with Dimmendaal’s (2001) much smaller survey of ATR, showing an ATR-less zone in both the Mande-sphere, and also in Central Africa.

In our present survey, we aim to build on this previous areal-typological work with the following goals: provide a more precise and nuanced definition of ATR contrast and harmony, and include significantly more language varieties to demarcate more precise macro-isoglosses. We can compare our approach to ATR typology with that of the classification in Casali (2003, 2008). Casali’s work shows three types of vowel inventories with [ATR] shown in (3). 5Ht systems have a full set of ATR counterparts in high and mid vowels (ex. a.). In contrast 4Ht systems show a gap for one height, either among high [-ATR] vowels (b.), or mid [+ATR] vowels (c.).

(3) Vowel inventory types with [ATR] (Casali 2003:308-9)

|                |                |                |
|----------------|----------------|----------------|
| a. 5Ht system  | b. 4Ht(M)      | c. 4Ht(H)      |
| i            u | i            u | i            u |
| ɪ            ʊ |                | ɪ            ʊ |
| e (ə)    o     | e (ə)    o     | (ə)            |
| ɛ            ɔ | ɛ            ɔ | ɛ            ɔ |
| a              | a              | a              |

We classify 5Ht systems as the most restricted type of ATR, which we classify in our database as ‘Strict ATR systems’. Under this definition, ATR harmony must demonstrate cross-height harmony, i.e. [+ATR] high vowels only occur with [+ATR] mid vowels within roots and across some morpheme boundary. Therefore, Strict ATR systems have 9 or more vowels with high and mid counterparts, showing clear cross-height harmony both in static patterns and dynamic patterns. The Degema examples in (2) above exemplify this type.

We also include in this strict type those languages which show cross-height harmony only at the surface phonetic level. In such systems, at the phonological level their phoneme contrasts resemble the 4Ht systems in (3b-c) above, but one of these phonemes has both [+ATR] and [-ATR] allophones depending on the ATR context. For example, Kakwa [keo] (Nilotic: South Sudan – Onziga & Gilley 2012) has an ATR harmony system with the phonemes /i ɪ ɛ a ɔ ʊ u/ without mid [+ATR] \*/e o/. A sample of words with identical vowels is provided in Table 2.

| /i/     | /ɪ/    | */e/ | /ɛ/          | /a/     | /ɔ/   | */o/ | /ʊ/    | /u/    |
|---------|--------|------|--------------|---------|-------|------|--------|--------|
| pirî    | wírî   | -    | léme         | máta    | mókɔ  | -    | mógó   | pútú   |
| ‘place’ | ‘song’ |      | ‘grass type’ | ‘chief’ | ‘leg’ |      | ‘body’ | ‘dust’ |

Table 2: Kakwa vowel phonemes

However, when /ε ɔ/ appear in the context of [+ATR] /i u/, they surface as [+ATR] variants [e o] respectively, e.g. /píré/ ‘fatten’ [píré] and /ópú/ ‘corpse’ [ópú]. We classify such systems as ‘Strict ATR’ because they show cross-height harmony as well as dynamic patterns.

For many of Casali’s 4Ht systems – especially 4Ht(M) – we code these systems as ‘Trace ATR systems’ (also called “incomplete” harmony systems – Ladefoged 1964:37). In these cases, there are vowel restrictions which resemble ATR, but do not demonstrate dynamic cross-height harmony. For example, most show a restriction in the mid-heights where mid-close vowels do not co-occur with mid-open and vice versa, i.e. \*[e...ε] and \*[ɔ...o] in Table 3.

| Sequences |             | Sequences |             |
|-----------|-------------|-----------|-------------|
| √         | i...e i...o | √         | u...e u...o |
| √         | i...ε i...ɔ | √         | u...ε u...ɔ |
| *         | e...ε e...ɔ | *         | o...ε o...ɔ |
| *         | ε...e ε...o | *         | ɔ...e ɔ...o |

Table 3: Trace systems’ mid height co-occurrence restrictions

Such systems may show static or dynamic restrictions in the mid-heights. For example, although Gbeya [gya] (Gbaya: CAR) has a static restriction in roots as in the table above, Samarin (1966:50) notes that the demonstrative suffix /-ε/ does not alternate in the context of /e o/, e.g. [ə a yór-ε] ‘there he stands nearby’. Even in those systems which show dynamic co-occurrence restrictions in mid-heights resulting in allomorphy, if there is no clear cross-height harmony such a system was coded as Trace ATR.

Finally, those languages which do not have cross-height ATR harmony and have no trace features of ATR harmony are classified as ‘No ATR’. We summarize the three patterns in (4). This shows that approximately half the language varieties surveyed in this area have a Strict/Trace ATR system, and half lack ATR altogether. The map in Figure 2 shows the geographic distribution of these types.<sup>2</sup>

(4) Coding for ATR harmony (n = 615)

|            |   |     |     |
|------------|---|-----|-----|
| Strict ATR | Has cross-height harmony with dynamic patterns  | 193 | 31% |
| Trace ATR  | No cross-height harmony but with trace patterns | 119 | 19% |
| No ATR     | No trace patterns of an ATR system              | 303 | 49% |

<sup>2</sup> Note that although Trace ATR systems may descend from former Strict ATR systems (and display true traces of ATR), this is not necessarily the case and we make no claim as to the origin of these patterns by using the term ‘trace’. For example, Edo [bin] (Edoid: Nigeria – Elugbe 1989) descends from Proto-Edoid which uncontroversially had cross-height ATR harmony, but it is classified in our survey as No ATR because it no longer even has mid-height restrictions (e.g. its name /édó/). Moreover, Gbaya languages are mostly classified as Trace ATR, even though Proto-Gbaya is not constructed with cross-height Strict ATR harmony and no modern language has such a system either (Moñino 1995). Strictly speaking, our classifications are sensitive only to synchronic patterns and not diachronic history.

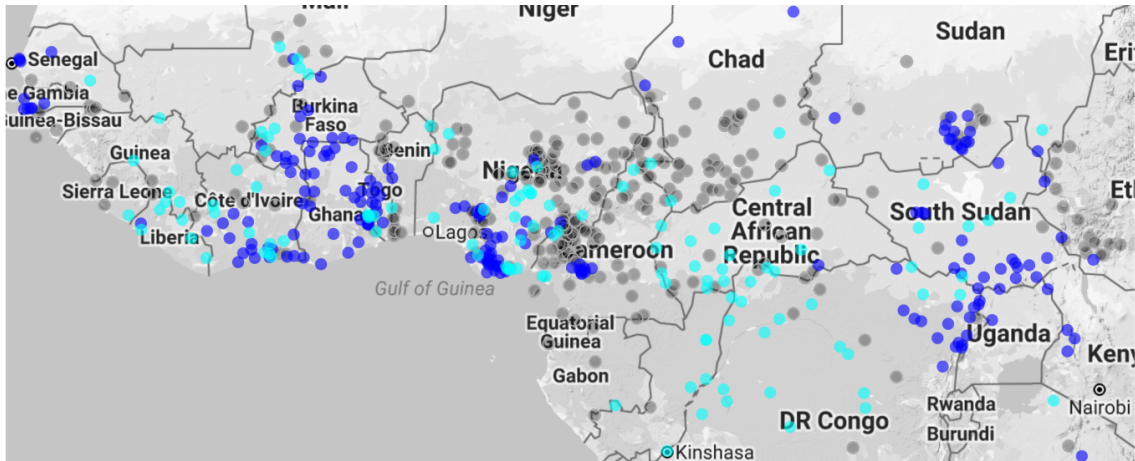


Figure 2: Geographic distribution of ATR Harmony

From this map, we can see that Strict ATR languages cluster in several areas. We term the first the Atlantic ATR Zone, which is a small zone of North Atlantic languages in Senegal and The Gambia. We term the second the West African ATR Zone, which encompasses a large portion of languages near the Gulf of Guinea stretching from Côte D'Ivoire to Nigeria along the coast and moving into Northern Ghana and Burkina Faso. The third is the East African ATR Zone, a large zone stretching from Kordofanian languages in the Nuba Mountains of southern Sudan extending south to Uganda/Kenya. Four outliers from these zones include [1] the Mbam languages (Bantu A) of central Cameroon (the cluster of dark blue dots surrounded by grey dots), [2] the Saharan languages of northern Chad (which might be interpreted as a distant extension of the East African ATR Zone), and the only two languages which do not appear on this map, [3] the Bantu language Malila [mgq] in southern Tanzania to the south of the East African ATR Zone, and [4] the Cushitic language Somali to the east of this zone.

Trace ATR systems are found in three main zones. We term the first the West African Trace ATR Zone, which is found at the Liberia/Cote D'Ivoire/Guinea/Burkina Faso confluence, and is composed primarily of Mande languages. The second is the Nigerian Trace ATR Zone, found primarily in Eastern and Southern Nigeria (on both sides of Strict ATR systems), and is not genetically homogenous. The third is the Central African Trace ATR Zone which extends from mainly Gbaya languages of the Central African Republic to Nilotic languages in South Sudan and Bantu C languages in the Democratic Republic of the Congo.

Finally, systems classified as No ATR are found widely through this area, and extend to all locations in Africa not on this map (Afro-Asiatic languages to the North and East, and Bantu/'Khoisan' to the South). We highlight two places in particular. One is the West African ATR-less Zone found in Guinea-Bissau/Guinea/Sierra Leone/Mali between the Atlantic ATR Zone and the West African Trace ATR Zone. The second is the much larger Central African ATR-less Zone which extends in the north from southern Chad and Sudan moving diagonally to northern and eastern Nigeria, and much of Cameroon, Equatorial Guinea, and Gabon. The large latter zone is genetically heterogeneous, being composed of Niger-Congo families (e.g. Grassfields Bantoid, Bantu A, East Benue-Congo, Kainji, Platoid, Jukunoid, Adamawa, Ubangi), Nilo-Saharan families (e.g. Central Sudanic), and Afro-Asiatic families (all four Chadic sub-groups).

2.2. SURVEY OF INTERIOR VOWELS. The second feature of African vowel systems which we surveyed was the presence of interior vowels. Interior vowels are defined as vowel qualities within the interior regions of the vowel space as characterized by the International Phonetic Alphabet. These include front rounded vowels, all non-low central vowels, and unrounded non-low back vowels. These are summarized in (5).

- (5) Vowel quality distinctions
- a. Interior vowels [y ʏ ø œ i ɨ u ə ɘ ə ɜ ɛ ɞ ɯ ʊ ʁ ʌ]
  - b. Peripheral vowels [i ɪ e ɛ æ a u ʊ o ɔ ɑ ɒ]

Interior vowels are prevalent throughout Central Africa, a fact known by language family specialists of the area (Bantoid – Watters 1989:414; Chadic – Gravina 2014:147). Often, languages of this region have more than one interior vowel, e.g. high and mid vowels /i u ə/ found in minimal pairs in Kejom [bbk] (Grassfields: Cameroon – Akumbu & Fogwe 2012):

- (6) Kejom minimal pair with interior vowels /i u ə/
- |                  |                   |              |                        |
|------------------|-------------------|--------------|------------------------|
| i tʃɪ ‘in-law’   | i tʃɪ ‘fireplace’ | u tʃú ‘spit’ | u kəntʃù ‘wildcat sp.’ |
| e tʃê ‘minimize’ | ə tʃǒ ‘kick’      | o tʃô ‘pass’ |                        |

Further, to our knowledge only Thomas et al. (1973) make a claim that interior vowels are an areal feature of Central Africa, and explicitly connect families of this area of different phyla.

From our sample of 615 African language varieties, we surveyed and coded for the presence of interior vowels split into four categories. The first is Present – Phonemic, and consists of languages where one or more interior vowels are analyzed as phonemic and not derivable from phonological context, as in the Kejom example above in (6). The second is Present – Non-phonemic, and consists of languages which exhibit allophonic interior vowels where a peripheral vowel is realized as an interior vowel in some phonological context, e.g. /i/ is realized as [i] in closed syllables in Horom [hoe] (Platoid: Nigeria) which otherwise does not have interior vowels (Nettle 1998). This encoding also includes languages in which the only interior vowel was epenthetic or a reduced variant of all (or most) vowel qualities. A third encoding consists of languages in which the only phonemic interior vowel is the [+ATR], [+low] vowel /a/ which is counterpart to [-ATR] /a/. This [+ATR] ‘low’ vowel is often transcribed with /ə/ or /ɜ/, and often acoustically resembles them, and therefore under our strict definition of interiority qualifies as an interior vowel rather than a peripheral vowel. In order not to conflate languages where /ə/ operates in an [ATR] system and those where /ə/ which does not, we overtly code for this distinction. Most ATR harmony languages which have a complete set of [+ATR]/[-ATR] counterparts are coded with this value, as in the Degema examples in (1) above.

Finally, those languages which provide no positive evidence for any phonemic or allophonic interior vowel are classified as ‘Absent’ with respect to interior vowels. The map in Figure 3 presents the geographic distribution of interiority, which shows that approximately half of the languages surveyed had at least one phonemic or allophonic interior vowel of some sort, and half did not.



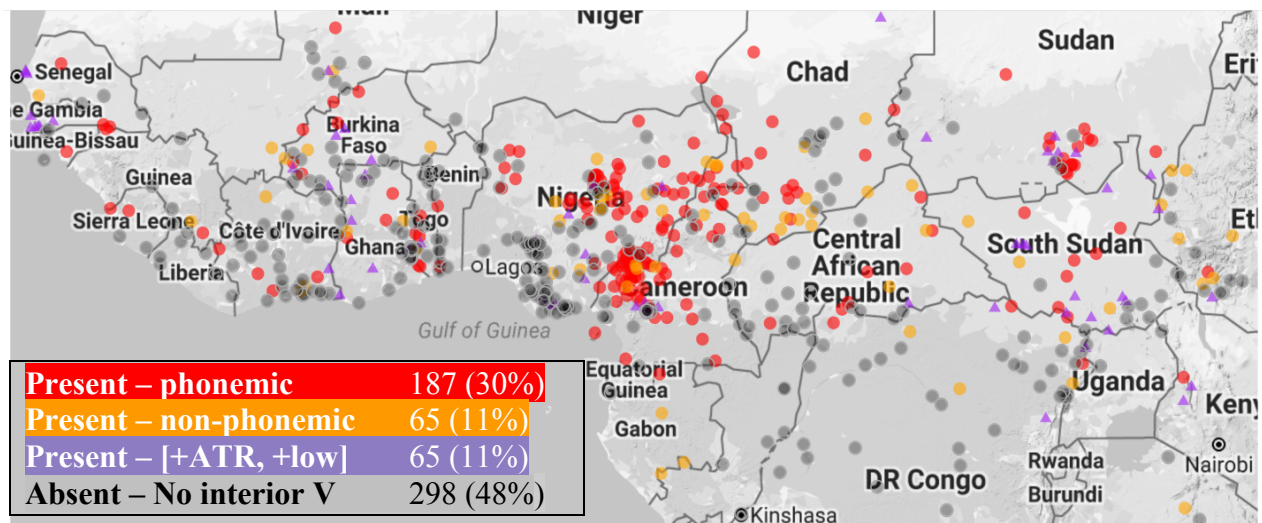


Figure 3: Geographic distribution of interior vowels

From this map, we see that the largest concentration of phonemic interior vowels (red dots) is found in what we term the Central African Interior Vowel Zone. This zone encompasses languages of western and southern Chad into eastern and northern Nigeria and a large part of Cameroon including much of the Grassfields. This area has one of the highest concentrations of language varieties in Africa, and is a confluence area of the three major phyla of Africa, namely Niger-Congo, Afro-Asiatic, and Nilo-Saharan. As such, this interior vowel zone is genetically heterogeneous, including language families East Benue-Congo (e.g. Grassfields Bantu), Adamawa, Ubangian, Kainji, Platoid, Jukunoid, Biu-Mandara Chadic, West Chadic, Masa Chadic, and Sara-Bongo-Bagirmi.

A second concentration of red dots indicating phonemic interior vowels can be seen in the East African Interior Vowel Zone, from the Nuba Mountains of southern Sudan to Nilotic languages of South Sudan. In addition, there are numerous red dots which appear sporadically from Senegal through Guinea, Cote D’Ivoire, Burkina Faso, Togo, and Benin, and in the Central African Republic, illustrating that this phonological feature is less constrained to ‘tight’ geographic distribution than ATR.

Further, those languages which have non-phonemic interior vowels (orange dots) are largely concentrated in a band extending from central Nigeria east into Chad and the Central African Republic, and further east into South Sudan and Ethiopia. This is significant in that it connects the two Phonemic Interior Vowel Zones (though note the many grey dots in this area indicating lack of any interiority, discussed below). Additional non-phonemic areas include parts of the Mali/Burkina Faso/Cote D’Ivoire confluence, Gabon, and western Ethiopia.

Those languages whose only phonemic interior vowel was [+ATR] /ɔ/ (counterpart to /ɑ/) are found in three main regions, but do not occur in a concentration great enough to warrant a uniquely named areal zone. The first is the Atlantic region in Senegal and the Gambia. The second is in an area stretching from north to south from Burkina Faso to southern Ghana. The third is an area in East Africa in Sudan, South Sudan, and their borderlands, intermixed with phonemic and non-phonemic interior vowel languages. Additional pockets are found in southern, central, and eastern Nigeria, and central Cameroon (the latter consisting entirely of the Mbam languages mentioned above).

Finally, many areas lack interior vowels altogether. The most significant is the West African Interior Vowel-less Zone, being the large concentration of grey circles found extending from Sierra Leone and Guinea in the west to central Nigeria in the east. Notice however that within this zone, there are a number of languages which do have interior vowels of some type, though they clearly are the minority. Additionally, the Gbaya languages of western Central African Republic and the Bantu languages of the Democratic Republic of the Congo consistently lack interior vowels (Bantu, of course, spreads southwardly to the end of the continent)<sup>3</sup>. Further pockets include central Mali, central and western Chad (the Eastern Chadic family) and many languages at the South Sudan/Ethiopia/Kenya confluence.

**3. ATR harmony with respect to interior vowels.** Having established the distributions of ATR and interiority, we seek to establish their relationship by plotting the geographic distribution of their co-occurrence. Impressionistically, many ATR languages do not have interior vowels (other than [+ATR] /a/) and many languages with phonemic interior vowels do not have ATR. However, their co-occurrence is not impossible, and several languages in Africa show elaborate vowel systems with full sets of [ATR] contrasts for front, central, and back vowels. Examples below come from three genetically unrelated languages of considerable geographic distance, Guiberoua Bété [bet] (Kru: Côte d’Ivoire – Marchese 1983), Kanembu [kbl] (Saharan: Chad – Jouannet 1982), and Tima [tms] (‘Kordofanian’/Katla-Tima: Sudan – Bashir 2013).

| ATR | Front | Central | Back |
|-----|-------|---------|------|
| +   | i     | ɨ       | u    |
| -   | ɪ     | ɘ       | ʊ    |
| +   | e     | ə       | o    |
| -   | ɛ     | ʌ       | ɔ    |
| ∅   |       | a       |      |

Table 4: Guiberoua Bété

| ATR | Front | Central | Back |
|-----|-------|---------|------|
| +   | i     | ɨ       | u    |
| -   | ɪ     | ɘ       | ʊ    |
| +   | e     | ə       | o    |
| -   | ɛ     | ʌ       | ɔ    |
| ∅   |       | a       |      |

Table 5: Kanembu

| ATR | Front | Central | Back |
|-----|-------|---------|------|
| +   | i     | ɨ       | u    |
| -   | ɪ     | ɘ       | ʊ    |
| +   | e     | ʌ       | o    |
| -   | ɛ     | a       | ɔ    |

Table 6: Tima

In order to address this issue, we mapped these features against each other as in the maps below. The first map in Figure 4 presents the distribution of ‘strict’ interiority (only phonemic interior vowels) with respect to ATR, showing four values: purple dots indicate those systems with both ATR and phonemic interior vowels, blue dots are those with phonemic interior vowels but no ATR, red dots are those with ATR but no phonemic interior vowels, and grey dots are those languages with neither. The second map in Figure 5 presents ‘liberal’ interiority by ATR. Here, a language is classified as having interiority whether the vowel is phonemic or allophonic. In both maps, ATR only refers to those systems which were coded above as Strict ATR; those languages which were Trace ATR were collapsed together with No ATR. Those systems where the only interior vowel is [+ATR] /a/ are a special case and therefore are denoted with a triangle around the dot.

<sup>3</sup> Notable exceptions are the Northwestern Gbaya dialects, which have allophonic front rounded vowels (Moñino 1995:58–68).

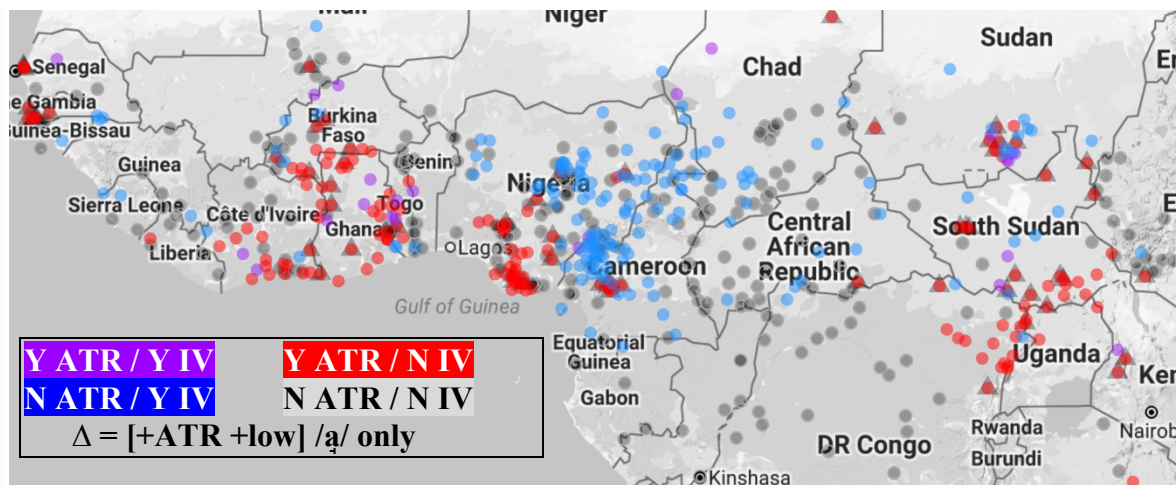


Figure 4: Distribution of ‘strict’ interiority by ATR

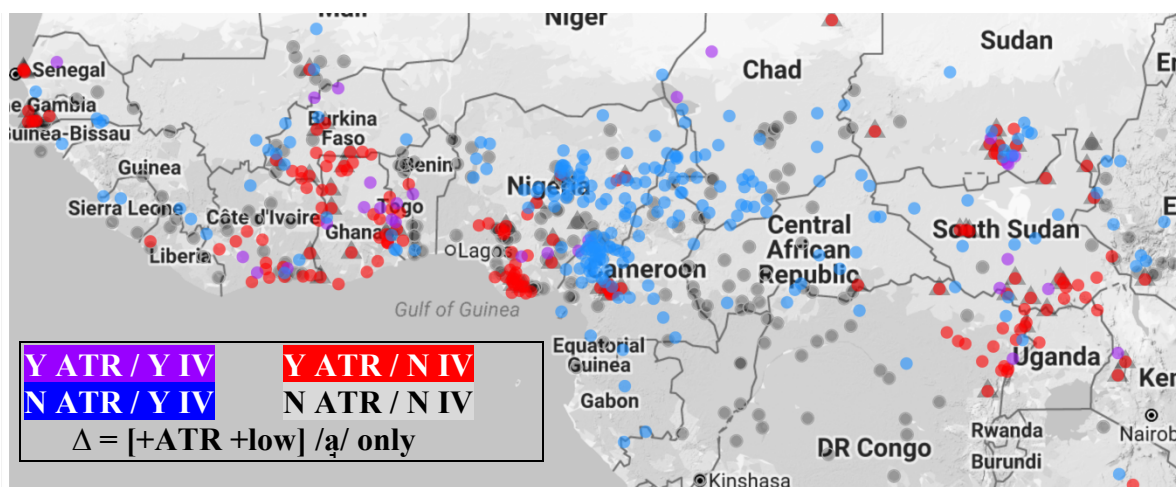


Figure 5: Distribution of ‘liberal’ interiority by ATR

We can see a number of trends from these maps. First, perhaps the most striking aspect is the large concentration of blue dots indicating the absence of ATR and the presence of interior vowels located in Central Africa. This corresponds to the Central African Interior Vowel Zone. This is shown both in the ‘strict’ interpretation of interiority in Figure 4 as well as the ‘liberal’ one in Figure 5. Of those languages in this area which are not marked as blue, the vast majority are grey (N ATR / N IV), indicating that a vast area of Central Africa categorically lacks cross-height ATR altogether (supporting earlier statements in Boyd 1989:197 and Dimmendaal 2001:370). These distributions reveal two quite distinct vowel profiles within the Macro-Sudan Belt spread across large sub-areas.

This fact is important given that ATR harmony is a linguistic criterion defining the Macro-Sudan Belt in Güldemann (2008), with the ‘hottest’ concentration of such criteria centered around Cameroon/Central African Republic (Güldemann 2011:110). Our maps show that the West and East African ATR Zones are entirely disconnected and should therefore not be conflated as belonging to the same areal zone strictly speaking. We make two additional points. First, the large number of Trace ATR systems between these two ATR zones may in fact act as an areal ‘bridge’, an idea which needs to be explored. Second, we fully acknowledge that the African macro-areas put forward in Güldemann (2008) and Clements & Rialland (2008) are not

‘sprachbunds’ in the traditional sense of unrelated language varieties systematically converging on identical linguistic structure, but rather are linguistic zones within which features can spread more easily than to areas which are not in the same zone (Güldemann p.c.). We agree that there is sufficient evidence for some version of a Macro-Sudan Belt to require some sort of historical explanation, especially given the rareness of ATR harmony outside of Africa.

Although less geographically categorical, there are two large concentrations of red dots indicating the presence of ATR and the absence of interiority, corresponding to the West African ATR Zone and the East African ATR Zone respectively. Moreover, we see pockets of purple dots where systems have both ATR and phonemic interior vowels, but no strong areal signal of this type emerges. These include pockets in Senegal, southern Cote D’Ivoire (Table 4 above), the Ghana/Togo borderlands, southeastern Nigeria, western Chad (Table 5 above), and scattered in the East African ATR Zone (Table 6 above). Visualizing the distribution of ATR by interiority highlights the fact that these two types of vowel contrasts rarely occur in the same phonological system. We therefore interpret these two types of contrast as inversely related.

Pearson’s Chi-squared ( $\chi^2$ ) tests of independence were performed to examine the relationship between presence of ATR harmony and presence of interior vowels in the ALFA sample. Table 7 shows four versions of the test, using either the ‘strict’ or ‘liberal’ interpretations of interiority, as well as a ‘strict’ interpretation of ATR where only Strict ATR systems with dynamic cross-height harmony count as having ATR, and a ‘liberal’ interpretation of ATR where Strict ATR and Trace ATR are collapsed together as ‘Y ATR’. All four of the resulting tests suggest that the variables ATR and IV are not independent regardless of how they are defined ( $p < 0.001$ ).<sup>4</sup> Our results can be interpreted as showing that elaboration of ATR harmony contrasting along the acoustic dimension of F1 is antagonistic to elaboration of interiority along the dimension of F2.

|   |     |     |        |       |
|---|-----|-----|--------|-------|
| <b>Strict ATR/<br/>Strict Interiority</b>   |     |     | Yes IV | No IV |
| Yes ATR                                     | 24  | 169 |        |       |
| No ATR                                      | 163 | 258 |        |       |
| $\chi^2 = 41.699, p < 0.001$                |     |     |        |       |
| <b>Strict ATR/<br/>Liberal Interiority</b>  |     |     | Yes IV | No IV |
| Yes ATR                                     | 30  | 163 |        |       |
| No ATR                                      | 222 | 200 |        |       |
| $\chi^2 = 73.691, p < 0.001$                |     |     |        |       |
| <b>Liberal ATR/<br/>Strict Interiority</b>  |     |     | Yes IV | No IV |
| Yes ATR                                     | 45  | 267 |        |       |
| No ATR                                      | 142 | 161 |        |       |
| $\chi^2 = 74.927, p < 0.001$                |     |     |        |       |
| <b>Liberal ATR/<br/>Liberal Interiority</b> |     |     | Yes IV | No IV |
| Yes ATR                                     | 61  | 251 |        |       |
| No ATR                                      | 191 | 112 |        |       |
| $\chi^2 = 118.39, p < 0.001$                |     |     |        |       |

Table 7:  $\chi^2$  tests showing statistical significance

**4. Concluding remarks.** In this paper, we investigated the distribution of vowel systems in the Macro-Sudan Belt using a survey of 615 language varieties coding for phonological features related to Advanced Tongue Root (ATR) harmony compared to the presence of interior vowels. Our results reveal an Atlantic ATR Zone, a West African ATR Zone, and an East African ATR Zone, which are geographically unconnected to one another. We show that between the West and East Zones is an extensive, genetically heterogeneous region of Central Africa where ATR is

<sup>4</sup> We should note that these tests are not sensitive to genetic affiliation. As our stated goal was exhaustive coverage of these areas, our database is not a genetically balanced sample (see Table 1).

systematically absent, which we call the Central African ATR-less Zone. Further, we demonstrate the existence of a Central African Interior Vowel Zone which overlaps significantly with the ATR-less Zone, suggesting that ATR and interiority have an antagonistic relationship in this region. Chi-squared tests reveal a significant relationship between ATR and interiority, supporting their mutual exclusiveness as defining two different vowel system profiles within the Macro-Sudan Belt macro-area.

Two further directions of this research are [1] correlating these linguistic areas with socio-historical events which could plausibly explain the formation of these areas and in particular the widespread macro-distribution of ATR despite the Central African ATR-less Zone, and [2] the loss and gain of ATR and interior vowels in the diachrony of African languages. Some plausible diachronic pathways are given below:

|     |                               |           |             |         |
|-----|-------------------------------|-----------|-------------|---------|
| (7) | Plausible diachronic pathways |           |             |         |
| a.  | ATR loss before interior gain | Stage 1   | Stage 2     | Stage 3 |
|     |                               | /i ɪ u ʊ/ | /i u/       | /i ɪ u/ |
| b.  | Interior gain before ATR loss | Stage 1   | Stage 2     | Stage 3 |
|     |                               | /i ɪ u ʊ/ | /i ɪ ɛ u ʊ/ | /i ɪ u/ |
| c.  | ATR directly to interior      | Stage 1   | Stage 2     |         |
|     |                               | /i ɪ u ʊ/ | /i ɪ u ɯ/   |         |

One language family in particular which may be fruitful for these efforts is Central-Sudanic (Nilo-Saharan phylum). Of the 36 Central-Sudanic languages we surveyed, roughly half are Strict ATR (n = 16) and half No ATR (n = 18), with a similar even distribution with respect to interiority types. Proto Central-Sudanic was probably spoken in at the DRC/Uganda/South-Sudan confluence, i.e. in the East ATR zone (Ehret 1974, a.o.). The Central-Sudanic branch Proto Sara-Bongo-Bagirmi (SBB) was probably spoken around the same area between South Sudan and the Central African Republic, and from there SBB speakers migrated westward outside of the East ATR zone and into the Central African Interior Vowel Zone in southern Chad and northern Central African Republic (Boyeldieu 2006, 2009). Those Central Sudanic and SBB languages still spoken in the East ATR zone have ATR. However, SBB languages which moved westwardly into the Central African Interior Zone do not have ATR distinctions, and rather have developed interior vowels, e.g. all Sara languages except the Sara Kaba subgroup which is also the easternmost (Keegan 1995, 2013).<sup>5</sup> This is very suggestive of a scenario in which SBB languages adapted their phonological profile to the new area they were entering (i.e. lost their ATR contrast and harmony and gained interior vowels). Further investigations of African language histories will be revealing for the development and spread of these vowel properties across the Macro-Sudan Belt.

## References

[references to the primary sources used in our database are available by request]

Akumbu, Pius, & Evelyn Fogwe. 2012. *A pedagogic grammar of Babanki*. Köln: Rüdiger Köppe Verlag.

<sup>5</sup> Boyeldieu (2006, a.o.) remains agnostic as to whether proto-SBB is to be reconstructed with ATR, but does not consider it an unlikely hypothesis (p.c.).

- Bashir, Abeer. 2013. Conditions on feature specification in Tima. In Thilo Schadeberg & Roger Blench (eds.), *Nuba Mountain Language Studies*. 269-280. Köln: Köppe.
- Boyd, Raymond. 1989. Adamawa-Ubangi. In John Bendor-Samuel & Rhonda L. Hartell (eds.) *The Niger-Congo languages*. 178–215. Lanham, Maryland: University Press of America.
- Boyeldieu, Pascal. 2006. Présentation des langues sara-bongo-baguirmiennes. Unpublished oral presentation, Paris, CNRS-LLACAN. Available online at <http://sumale.vjf.cnrs.fr/SBB/presentation.html> (last consulted: 17 April 2017)
- Boyeldieu, Pascal. 2009. Entre lac Albert et lac Tchad : l'expansion sara-bongo-baguirmiennne. In H. Tourneux & N. Woïn (eds.), *Migrations et mobilité dans le bassin du Lac Tchad. Actes du 13e colloque international du Réseau Méga-Tchad (Maroua, 31 oct.-3 nov. 2005)*. 239-252. Marseille: IRD Editions.
- Casali, Roderic F. 2003. [ATR] value asymmetries and underlying vowel inventory structure in Niger-Congo and Nilo-Saharan. *Linguistic Typology* 7. 307-382.
- Casali, Roderic F. 2008. ATR Harmony in African Languages. *Language and Linguistics Compass* 2. 496-549.
- Clements, George N. & Anne Rialland. 2008. Africa as a phonological area. In Bernd Heine & Derek Nurse (eds.), *A linguistic geography of Africa*. 36–85. Cambridge: Cambridge University Press.
- Dimmendaal, Gerrit. 2001. Areal Diffusion versus genetic inheritance: An African perspective. In R.W. Dixon & A. Aikhenvald (eds.), *Areal diffusion and genetic inheritance*. 358-392. Oxford: Oxford University Press.
- Donohue, Mark, Rebecca Hetherington, James McElvenny & Virginia Dawson. 2013. World phonotactics database. Department of Linguistics, The Australian National University. <http://phonotactics.anu.edu.au>. Accessed 2015-04-13.
- Ehret, Christopher. 1974. Some thoughts on the early history of the Nile-Congo watershed. *Ufahamu: A Journal of African Studies* 5(2). 85–112.
- Elugbe, Ben O. 1989. *Comparative Edoid: Phonology and lexicon*. Port Harcourt: University of Port Harcourt Press.
- Gravina, Richard. 2014. *The phonology of Proto-Central Chadic*. Utrecht: LOT School.
- 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.
- Güldemann, Tom. 2011. Proto-Bantu and Proto-Niger-Congo: Macro-areal typology and linguistic reconstruction. In Osamu Hieda, Christa König & Hiroshi Nakagawa (eds.), *Geographical typology and linguistic areas*. 109-141. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Hammarström, Harald, Robert Forkel, Martin Haspelmath, & Sebastian Bank. 2016. *Glottolog* 2.7. Jena: Max Planck Institute for the Science of Human History. (Available online at <http://glottolog.org>. Accessed on 2017-03-22.)
- Hartell, Rhonda L. (ed.). 1993. *Alphabets de Langues Africaines*. Dakar: UNESCO and Société Internationale de Linguistique.
- Jouannet, Francis. 1982. *Le kanembou de Ngaldoukou, langue saharienne parlée sur les rives septentrionales du lac Tchad: phonématique et prosodie*. (Société d'études linguistiques et anthropologiques de France, 91-92.) Paris: Société des Etudes Linguistiques et Anthropologiques de France (SELAF).

- Ka, Omar. 1994. *Wolof phonology and morphology*. Lanham, Md./London: University Press of America.
- Kari, Ethelbert E. 2007. Vowel harmony in Degema, Nigeria. *African Study Monographs* 28(2). 87-97.
- Keegan, John M. 1995. Sara vowel systems. In M. Lionel Bender & Thomas J. Hinnebusch (eds.), *Proceedings of the Sixth International Nilo-Saharan Linguistics Conference 1995*. *Afrikanistische Arbeitspapiere* 45. 83–105. (Available online at <http://morkegbooks.com/Services/World/Languages/SaraBagirmi/pdfs/SaraVowels.pdf>)
- Keegan, John M. 1997. *A Reference Grammar of Mbay*. Lincom Studies in African Linguistics, 14. München: Lincom Europa.
- Keegan, John M. 2013. Notes on the classification of Sara Kaba languages. (ms. available online at <http://morkegbooks.com/Services/World/Languages/SaraBagirmi/pdfs/NotesOnSaraKabaClassification.pdf>)
- Kutsch Lojenga, Constance. 2009. Two types of vowel harmony in Malila (M.24). In Margarida Petter & Ronald Beline Mendes (eds.), *Proceedings of the Special World Congress of African Linguistics: Exploring the African language connection in the Americas, São Paulo, 2008*, 109–128. São Paulo: Humanitas.
- Ladefoged, Peter. 1968. *A phonetic study of West African languages: An auditory-instrumental survey*. Cambridge: Cambridge University Press.
- Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.
- Marchese, Lynell. 1983. *Atlas linguistique kru. Essai de typologie*. (3rd revised edition). Abidjan: ACCT-ILA.
- Moñino, Yves. 1995. *Le Proto-Gbaya: essai de linguistique comparative historique sur vingt-et-une langues d'Afrique centrale*. Vol. 20. Leuven: Peeters.
- Moran, Steven, Daniel McCloy & Richard Wright (eds.) 2014. PHOIBLE Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://phoible.org>, Accessed on 2017-03-13.)
- Nettle, Daniel. 1998. *The Fyem language of northern Nigeria*. Vol. 136. München: Lincom Europa.
- Onziga, Yuga Juma & Leoma Gilley. 2012. Phonology of Kakuwâ (Kakwa). *Occasional Papers in the Study of Sudanese Languages* 10. 1-15.
- Saeed, John. 1999. *Somali*. Amsterdam: John Benjamins.
- Samarin, William J. 1966. *The Gbeya language: Grammar, texts, and vocabularies*. Berkeley & Los Angeles: University of California Press.
- Stewart, J. M. 1967. Tongue root position in Akan vowel harmony. *Phonetica* 16. 185-204.
- Thomas, Jacqueline M.C., Luc Bouquiaux, Jean-Pierre Caprile, France Cloarec, Gladys Guarisma, Claude Hagége, Yves Moñino, Christiane Paulian, Paulette Roulon & Pierre-Francis Lacroix. 1973. Aires de phonèmes et aire de tons dans les langues d'Afrique Centrale. In J.M.C. Thomas & L. Bernot (eds.), *Langues et techniques. Nature et société*. Vol. 1. 111-121. Paris: Klincksieck.
- Velde, Mark van de. 2008. *A grammar of Eton*. Berlin: Walter de Gruyter.
- Watters, John. 1989. Bantoid overview. In John Bendor-Samuel & Rhonda L. Hartell (eds.), *The Niger-Congo languages*. 401-420. Lanham, Maryland: University Press of America.