West African languages

Linguistic theory and communication

Edited by
Nina Pawlak
Izabela Will

WYDAWNICTWA UNIWERSYTETU WARSZAWSKIEGO
WUW

West African languages

West African languages

Linguistic theory and communication

Edited by
Nina Pawlak
Izabela Will

WUW

Reviewers
Jadwiga Linde-Usiekniewicz
Anne Storch

Commissioning Editor
Ewa Wyszyńska

Proofreading
Bożena Lesiuk

Cover Design
Anna Gogolewska

Illustration on the Cover
siloto/shutterstock

Layout and Typesetting
Dariusz Górski

Published with financial support from the University of Warsaw
Published with financial support from the Faculty of Oriental Studies, University of Warsaw
Published with financial support from the University of Warsaw Foundation

Nina Pawlak ORCID 0000-0001-6949-3921
Izabela Will ORCID 0000-0002-2797-1160

ISBN 978-83-235-4623-8 (print)
ISBN 978-83-235-4631-3 (pdf online)
ISBN 978-83-235-4639-9 (e-pub)
ISBN 978-83-235-4647-4 (mobi)

Wydawnictwa Uniwersytetu Warszawskiego
00-838 Warszawa, ul. Prosta 69
E-mail: wuw@uw.edu.pl
Online bookstore: www.wuw.pl

1st Edition, Warsaw 2020

Print and binding
POZKAL

Contents

PART IV. SOCIAL AND COMMUNICATIVE ASPECTS OF LANGUAGES

General abbreviations and symbols

C	Consonant
H	High tone
L	Low tone
M	Mid tone
NP	Noun Phrase
POS	Part of Speech
PP	Prepositional Phrase
SVC	Serial verb construction
TAM	Tense-Aspect-Mood
WALS	*World Atlas of Language Structures*
V	Vowel

Standard glossing labels

1	first person
2	second person
3	third person
4	impersonal pronoun
ACC	accusative
ADJ	Adjective
ADV	adverb(ial)
AGR	agreement
AUX	auxiliary
COMP	complementizer
COMPL	completive
CONT	continuous
COP	copula
DAT	dative
DEF	definite
DEM	demonstrative
DET	determiner
F	feminine
FOC	focus
FUT	future
GEN	genitive
ICP	intransitive copy pronoun
IDEO	ideophone
IO	indirect object
IMP	imperative
IPFV	imperfective
LOC	locative
M	masculine
NEG	negation, negative
OBJ	object
PART	particle
PFV	perfective
PL	plural
PLUR	pluractional (verbal plural)
POSS	possessive
PRS	present
Q	question
REL	relative
SBJV	subjunctive
SG	singular

Gian Claudio Batic
ORCID 0000-0002-5040-7891
Università degli Studi di Napoli "L'Orientale"

Measuring phonological complexity in West African languages

Abstract

The discussion on the complexity of natural language is a fascinating topic that has been treated by scholars from different philosophical and theoretical perspectives. The main challenge to overcome when studying complexity is represented by its quantification: discussing complexity means dealing with objective measurements. Since languages are systems, i.e. they are made up of elements, it is possible to examine the structural complexity of a language by counting the elements present in the system. Systems (that is, languages) are in turn made of sub-systems (that is, areas), each sub-system being described by a series of features whose inventory sizes can be, for example, relatively small, relatively large, or average. This paper aims at formulating an *Index of Phonological Complexity* (IPC) based on the typological features covering the phonological area as defined in the *World Atlas of Language Structures* (WALS). After a brief discussion on these features, their recoding and subsequently their normalisation to a common scale will be argued for. Then, three indexes of phonological complexity will be proposed and applied to West African languages. Given its high degree of linguistic diversity determined by both genetic and typological variety, West Africa is an interesting ground for measuring complexity as well as a promising laboratory for further calibration and refinement of the indexes.

Keywords: phonology, complexity, West Africa

1. Complexity: what is it?

The idea that languages are somehow 'complex' has a relatively long history. In the 19th century, languages were seen as products of the communities that spoke them. The principle was clear: sophisticated communities used sophisticated languages and the most sophisticated communities were those represented by nations. Since the complexity of a given language was measured against abstractions such as the 'spirit of the nation', languages of nationless communities were deemed – here the lexical choice is strictly dependent on the context – simpler, primitive, or inferior. Races that built nations had something that nationless groups did not have: the ability to express abstractions (cf. Herder 1772; von Humboldt 1836). This (hard-to-die) idea started fading away, at least in the scientific community, by the second half of the following century, when the Academia shifted away from the fatally romantic assumptions of the past and replaced the old ideology with a new one: languages, it was said now, do not differ much from each other in terms of complexity. The over quoted passage of Charles Hockett according to which "it would seem that the total grammatical complexity of any language [...] is about the same of any

other" (Hockett 1958: 180) became emblematic of the new *Zeitgeist* permeating the view of linguistics on the differences between languages. The rise of the generative school in the 1960s neutralised the problem: the postulated existence of a universal grammar innate to *Homo sapiens* made all the observable differences between natural languages accidental and negligible. Saying that all languages are equally complex is, from a generativist point of view, true and beyond the point at the same time.

The topic of linguistic complexity, however, did not die out. Descriptions of single languages continued to be carried out and typology – also by virtue of its empirical foundation – became the right standpoint from which to observe what is going on in the realm of natural languages. More specifically, typology tells us what languages have and what they do not have, not only feature-wise but also in terms of inventories and their sizes.

The notion of linguistic complexity as an object of study in its own is newer. A significant number of publications appeared over the last decade and scholars from different theoretical and methodological frameworks have proposed a variety of approaches with the aim to address linguistic complexity as a meaningful field of enquiry (among others, Miestamo et al. 2008; Sampson et al. 2009; Baechler & Seiler 2016). These studies define complexity in a quite homogenous way, namely by operating a clear-cut distinction between *relative* and *absolute* complexity.

Relative complexity is close to the popular notion that sees complexity as mostly related to the categorization of natural languages into 'easy languages' vs. 'difficult languages', which translates into 'easy-to-learn languages' vs. 'difficult-to-learn languages'. The widespread idea according to which complex languages are difficult – i.e. complexity equals difficulty – hides a more pragmatic concern: how difficult is it for a speaker of language X to learn language Y? We are all familiar with statements like

German is easier to learn than French if your mother tongue is Dutch.
A Russian will not have so much trouble in mastering Serbian.
Oh, you are Italian? Perfect, so you understand Spanish too!

This notion of complexity is based on the perception people have of the language they speak (source language) and the language they want to learn (target language). This kind of comparison is essentially empirical and is built upon unsurprising similarities between source and target – or, better, upon similarities between homologous systems of source and target respectively (X phonology vs. Y phonology, X morphology vs. Y morphology, and so on). Rather unsurprisingly, when (in)direct experience does not support our stereotyped understanding of complexity (=when the operation of relativizing complexity falls short of data) we get lost:

Oh boy, she speaks Mekens![1] Tuvinian[2] will be no trouble at all.

Complexity can be understood from three different perspectives: cognitive, developmental, and absolute. Cognitive complexity relates to the processing costs attached to

[1] Tupian, Tuparic (South America), ISO 639-3 skf.
[2] Turkic, South Siberian Turkic, ISO 639-3 tyv.

linguistic structures, while developmental complexity to the way and the order of acquisition of such structures (Pallotti 2015: 117-118). These two kinds of complexities are relative to the speaker/user/learner.

Structural complexity (the kind of complexity we are interested in), on the contrary, deals with the 'absolute' and is defined independently of the speaker/user/learner. This notion of complexity is based on (a) the number of elements within a given system, and (b) the relations of these elements within the system. There is no correlation between 'absolute' complexity and 'relative' complexity: from an acquisitional point of view, it is perfectly normal to have structurally simple systems that are very hard to manage cognitively.

Complexity will be treated here as something that can be measured: the elements of different areas of the grammars can be 'counted'. In phonology, such notion of absolute or grammatical complexity translates into inventory sizes. An inventory with a large number of items will be labelled [+complex], whereas an inventory with a small number of items will be labelled [–complex]. This principle is exemplified by the vowel systems of Kushi (Afroasiatic, Chadic, Nigeria, ISO 639-3 kuh) and Diyari (Pama-Nyungan, Australia, ISO 639-3 dif): the vowel system of Kushi is more complex than that of Diyari.

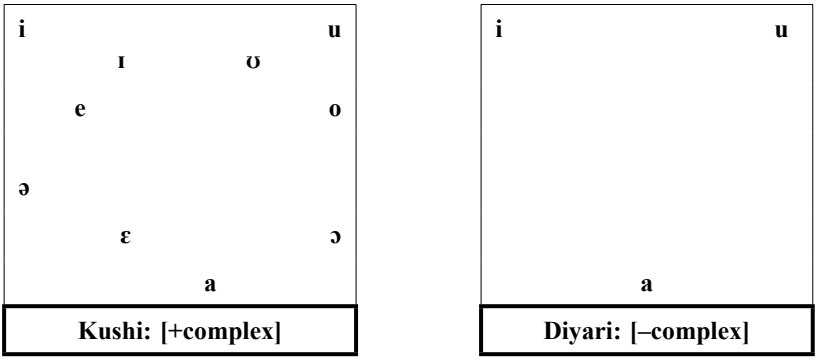

Table 1 – Vowel systems of Kushi (Afroasiatic, Chadic / Nigeria) and Diyari (Pama-Nyungan / Australia): [±complexity]

The current debate on the complexity of languages has been focusing on the definition and metrics of complexity, challenging Hockett's assumption on 'total complexity'. Once the idea that a language is a mechanism whose components can be analysed in terms of numbers (later on we will see what this exactly means) is accepted, then the master question becomes: 'how can we measure the global complexity of languages?'.

Let's step back to the 'total complexity' notion stated by Hockett. The notion that sees all languages as equally complex has a strong theoretical implication. If the overall complexity of a language X is equal to the overall complexity of a language Y, then there must a compensation effect in play between different areas of the grammar. Let's suppose, for example, that language X has a very simple phonology (e.g. 3 vowels, 17 consonant phonemes, no tones, and simple syllables such as V and CV). Then, according to the idea that all languages are equally complex, one would expect other areas of the grammar, such as morphology or syntax, to compensate for the poor phonological system. Analogously, if we took into account a language Y displaying a complex phonological system, then

we would expect other areas of the grammar to be simple or less complex. But the equality principle, reasonable (and politically-correct) it may seem, has two weak points. First of all, it is not true that all languages are equally complex: creoles, for example, are consistent in showing a 'simpler' grammar (Parkvall 2008). Secondly, the high complexity shown by certain languages in some areas of their grammars, e.g. verbal morphology, can't be equalled or compensated for by complex areas in other languages. In other words, there are cases where complexity is so high that any comparison based on the equality principle does not hold: some languages are simply more complex than others.

This paper builds on previous work on linguistic complexity (Bentz et al. 2016; Maddieson 2005; 2006; 2011) and addresses the possibility to measure phonological complexity from a typological standpoint, i.e. a global quantification specifying the phonological complexity of a cluster of typological features. This paper has two main purposes: the formulation of an *Index of Phonological Complexity* (IPC), and the application of such an index across West African languages. The present study is based on the typological features defined in the *World Atlas of Language Structures* (Dryer & Haspelmath 2013), namely the chapters on phonology (Maddieson 2013a-k; Anderson 2013; Goedemans & van der Hulst 2013a-d; and Hajek 2013). In order to define an IPC, I will shortly present the phonological features considered in the WALS. Then I will argue for a numerical recoding of these features and for their subsequent normalisation to a common scale. Having completed these preliminary steps, I will propose three indexes of phonological complexity and discuss their application to West African languages.

2. West Africa: a phonological zone and a testing ground

The rationale for the choice of West Africa as the testing ground for an index of phonological complexity is due to the extreme variety of the region in terms of phonological features. While an index is by definition universal (it is possible to apply it in any context to obtain an objective measurement), not all contexts in which an index can be used are equally interesting: the most productive way to test an index of phonological complexity would be its application in a context displaying a high density of languages belonging to different language phyla. The purpose of the index (and the comparative analysis of the results obtained through its application across the languages present in a given area) does not consist in highlighting the presence of this or that phonological feature, but rather in identifying the inter-dependence of features and how their inventory sizes increase or decrease when specific features coexist in the same language.

West Africa is defined here as the region bordered by the Atlantic Ocean to the west and the south, the Sahara desert to the north, and Cameroon/Chad to the east. The region represents only a portion of the 'Sudanic belt' – a vast phonological zone including West Africa and extending to Lake Albert and the Ethiopian-Eritrean highlands to the south-east and east respectively – identified by Clements & Rialland (2007). Clements & Rialland's subdivision of Africa in six phonological zones (North, East, Sudanic, Center, South, and Rift) is motivated by the observation that many phonological features specific to the African continent are geographically restricted, hence the necessity to define different zones on the basis of different clusters of phonological (i.e. segmental and prosodic) features. Almost

all the phonological features that characterise the Sudanic belt as a whole are also found in West Africa: if we accept the existence of a Sudanic belt (and certainly the evidence is compelling), then West Africa alone is an excellent representative of such phonological zone. The specificity of the West African/Sudanic region in terms of phonology is given by the presence of a) labial flaps, b) labial-velar stops, c) vowel and consonant nasality, d) ATR-based vowel harmony, and e) simple and complex tone systems.

Since West Africa is not treated here as a phonological zone *sensu stricto*, but rather as a meaningful testing ground where complexity can be measured and analysed (and, in a sense, observed in action) by virtue of a high degree of linguistic diversity, an important reason for basing a study on this region is the availability of data. Despite the fact that the scientific community is still lacking a thorough description of many African languages, the West African region, when it comes to phonological features, is fairly represented in the WALS and in print publications, thus allowing for a firm approach to phonology-based phenomena.

3. Phonological features

The WALS covers phonology by specifying 20 features:

#	WALS features: phonology	
1	1A	consonant inventories
2	2A	vowel quality inventories
3	3A	consonant-vowel ratio
4	4A	voicing in plosives and fricatives
5	5A	voicing and gaps in plosive systems
6	6A	uvular consonants
7	7A	glottalised consonants
8	8A	lateral consonants
9	9A	velar nasal
10	10A	vowel nasalisation
11	10B	nasal vowels in West Africa
12	11A	front rounded vowels
13	12A	syllabic structure
14	13A	tone
15	14A	fixed stress locations
16	15A	weight sensitive stress
17	16A	weight factors in weight-sensitive stress systems
18	17A	rhythm types
19	18A	absence of common consonants
20	19A	presence of uncommon consonants

Table 2 – WALS features for the phonological area

Each of these features displays a set of values. For a certain number of features the values are strictly related to the size of the reference inventory. This is the case of feature 1A (consonant inventories), 2A (vowel quality inventories), 12A (syllabic structure), and 13A (tone). The following tables illustrate the set of labels/values assigned to each of these features and the respective inventory size.

1A – WALS value	Number of consonants
Small	6-14
Moderately small	15-18
Average	22 (±3)
Moderately large	26-33
Large	34≤

Table 3 – Consonant: inventory types

2A – WALS value	Number of vowels
Small	2-4
Medium	5-6
Large	7-14

Table 4 – Vowels: inventory types

12A – WALS value	Syllable types
Simple	(C)V, CV
Moderately simple	CVC, CCV, CCVC
Complex	(C)(C)(C)V(C)(C)(C)(C)

Table 5 – Syllabic structures

13A – WALS value	Tone
No tone	–
Simple tone system	two-way contrast
Complex tone system	3≤

Table 6 – Tones

Rather than being tied to the size of an inventory, some features are defined on the basis of the presence (or absence) of a certain phonological element and in the way this element surfaces in the language. See for example feature 7A (glottalised consonants):

7A – WALS value
No glottalised consonants
Ejectives only
Implosives only

Glottalised resonants only
Ejectives and implosives
Ejectives and glottalised resonants
Implosives and glottalised resonants Ejectives, implosives, and glottalised resonants

Table 7 – Glottalised consonants

4. Measuring phonological complexity in a typological perspective

In order to obtain an index of phonological complexity, i.e. a value measuring the complexity of a cluster of features, it is necessary (1) to recode the values used in the WALS by converting them into a numerical scale, and (2) to normalise these new values to a common scale.

4.1. Values: recoding

The WALS categorises consonant, vowel, syllabic, and tone inventories according to their sizes and/or complexity. Consonant and vowel inventories are defined in terms of size, whereas syllabic structures and number of tones are referred to in terms of [± complexity]. We can operate a label normalisation and think the four systems in terms of structural complexity, i.e. we could range the size types of the four inventories on a scale that goes from [+simple, –complex] to [–simple, +complex], where a system's complexity and the number of its elements are directly proportional. Since the aim is to attain a measure of the phonological complexity in a given language with the ultimate goal to compare values, the labels will be recoded in terms of numeric values. In the tables below, numeric values have been assigned to the WALS inventory types – the lower the rank in terms of complexity, the lower the assigned value. The five types of consonant inventories have been arranged on a scale from 1 [+simple, –complex] to 5 [–simple, +complex].

WALS value	Recoded value
small	1
mod. small	2
average	3
mod. large	4
large	5

Table 8 – WALS values for consonant inventories coded on a [1, 5] scale

Values to vowel, syllabic, and tonal inventories have been assigned in the same way. These phonological systems display three major groups the inventories can be categorised into, hence the values have been assigned on a [1, 3] scale:

Vowels	Value	Syllables	Value	Tones	Value
small	1	simple	1	Ø	1
medium	2	mod. simple	2	simple	2
large	3	complex	3	complex	3

Table 9 – WALS values for vowel/syllable/tone inventories coded on a [1, 3] scale

Other features whose values have been rearranged on a [1, 3] scale are 4A (voicing in plosive and fricatives), 5A (voicing and gaps in plosive systems), and 6A (uvular consonants). For these features it is not possible to assign a different score to each of the intermediate values, as their degree of complexity is the same. Consider, for example, the case of feature 4A. The WALS defines 4 values, but not 4 different degrees of complexity. The absence of voicing contrast represents the minimum degree of complexity, the presence of contrast in plosive and fricative the maximum degree, but the plosives-only contrast and the fricatives-only contrast are undistinguishable in terms of complexity, hence they have been quantified with the same score. A four-value set has been coded to a three-value range.

4A – WALS values	WALS value [1, 3]
no voicing contrast	1
voicing contrast in plosive alone	2
voicing contrast in fricatives alone	2
voicing contrast in both plosives and fricatives	3

Table 10 – WALS values for plosive contrast coded on a [1, 3] scale

Similarly, the five-value feature 8A has been coded on a three-degree scale:

8A – WALS values	WALS value [1, 3]
no laterals	1
no /l/, but lateral obstruents	2
laterals, but no /l/, no obstruent laterals	2
/l/, no obstruent laterals	2
/l/ and lateral obstruent	3

Table 11 – WALS values for laterals [1, 3] scale

The seven-value feature 19A (presence of uncommon consonants) has been transformed into a binary system ([0, 1]):

19A – WALS values	WALS value [0, 1]
None	0
Pharyngeals and “th”	1
Pharyngeals	1

None	1
Labial-velars	1
Clicks, pharyngeals, and “th”	1
Clicks	1
“Th” sounds	1

Table 12 – WALS values for the presence of uncommon consonants on a binary [0, 1] scale

#	WALS feature		Scale	Type	Transformation
1	1A	consonant inventories	[1, 5]	ordinal	[0, 1]
2	2A	vowel quality inventories	[1, 3]	ordinal	[0, 1]
3	3A	consonant-vowel ratio	[1, 5]	ordinal	[0, 1]
4	4A	voicing in plosives and fricatives	–	mixed	[0, 1] + reordered
5	5A	voicing and gaps in plosive systems	–	mixed	[0, 1] + reordered
6	6A	uvular consonants	[1, 3]	mixed	[0, 1]
7	7A	glottalised consonants	–	mixed	[0, 1] binary
8	8A	lateral consonants	–	mixed	[0, 3] reordered
9	9A	velar nasal	–		[0, 1] binary
10	10A	vowel nasalisation	[1, 2]	binary	[0, 1] binary
11	10B	nasal vowels in West Africa	–	mixed	[0, 1] reordered
12	11A	front rounded vowels	–	mixed	[0, 1] reordered
13	12A	syllabic structure	[1, 3]	ordinal	[0, 1]
14	13A	tone	[1, 3]	ordinal	[0, 1]
15	14A	fixed stress locations	–	mixed	[0, 1] binary
16	15A	weight sensitive stress	–	mixed	[0, 1] binary
17	16A	weight factors in weight-sensitive stress systems	–	mixed	[0, 1] binary
18	17A	rhythm types	–	mixed	[0, 1] binary
19	18A	absence of common consonants	–	mixed	[0, 1] binary
20	19A	presence of uncommon consonants	[0, 1]	binary	[0, 1] binary

Table 13 – Transformations

4.2. Values: normalisation

In order to be comparable, the value ranges of the scales obtained through recoding ([1, 5], [1, 3] and [0, 1]) need to be normalised, i.e. to be transformed and arranged on a common scale. The three scales will be normalised to a [0, 1] interval. The reason behind the choice of the [0, 1] interval is that it will allow us to express the Index of Phonological Complexity as a fraction of 1, i.e. 0 followed by *n* digits. Moreover, the [0, 1] interval is consistent with the previous literature on linguistic complexity (Bentz et al. 2016) or diversity (Harmon & Loh 2010) indexes.

(1) $$Normalised(x_i) = \frac{x_i - X_{min}}{X_{max} - X_{min}}$$

Where:
X_{min} = the minimum value for variable X
X_{max} = the maximum value for variable X
x_i = the raw data to be normalised

As a result, the normalised values are as follows:

Recoded WALS values			New values
[1, 5]	[1, 3]	[0, 1]	[0, 1]
1	1	0	0
2			0.25
3	2		0.5
4			0.75
5	3	1	1.0

Table 14 – Normalised values: correspondences

5. Index of phonological complexity

An IPC can be obtained by calculating the feature value average per language:

(2) $$\text{IPC}_{WALS(20)} = \frac{\sum_{i=1}^{n} f_i}{n}$$

Where:
f_i = value of feature i
n = number of features available per language

The formula takes into account only the features for which the WALS assigns a value. Table A in the Appendix lists the West African languages present in the WALS according to their IPC (in descending order, where the last column indicates the number of features available).

5.1. Adjusting the index

The IPC based on the 20 phonological features specified in the WALS poses some questions. First of all, as shown in the table A of the Appendix, for many features there are no values. Then, it is reasonable to say that a complete picture of the values of phonological features is – at least with the quality and amount of data available at this moment – very difficult to attain. These two immediate observations lead to the following question: how many features do we have to take into account in order to calculate a reliable IPC? I will discuss three main options: (a) a comprehensive index, (b) a quantitative-qualitative index

based on the current availability of data, and (c) a 'quick' index based on a very limited number of features.

Comprehensive index based on 20 features:

The most complete index we can envisage is an index that takes into consideration all the phonological features. This index, the $IPC_{WALS(20)}$, is calculated using the formula indicated below.

(3) $$IPC_{WALS(20)} = \frac{\sum_{i=1}^{n} f_i}{n} \qquad \text{(where } 1 \leq n \leq 20\text{)}$$

The $IPC_{WALS(20)}$ is certainly comprehensive, but its extent seems to collide with the paucity of data in some phonological sub-areas, e.g. features such as 14A, 15A, 16A and 17A are rarely described and no value can be assigned to them. Moreover, this index considers both quantitative and qualitative features, whereas only strictly quantitative features are indicative of grammatical complexity.

Quantitative-qualitative index based on 13 features:

The $IPC_{WALS(13)}$ mirrors the data available in the WALS. This index leaves aside the features that are described only sporadically, considering instead the feature values that are statistically consistent in terms of presence in the WALS. Out of 20 features, we observe that 13 are those that make it to the WALS: 1A, 2A, 3A, 4A, 5A, 6A, 7A, 8A, 11A, 12A, 13A, 18A, 19A. While the $IPC_{WALS(13)}$ considers a relatively high number of features, it stills mixes quantitative and qualitative features.

(4) $$IPC_{WALS(13)} = \frac{\sum_{i=1}^{n} f_i}{n} \qquad \text{(where } n = 13\text{)}$$

Quantitative index based on 4 features:

The last index is perhaps the most intuitive. It cuts the issue down to size by considering only 4 features: consonant inventory (1A), vowel inventory (2A), syllabic structure (12A), and tonal system (13A).

(5) $$IPC_{WALS(4)} = \frac{\sum_{i=1}^{n} f_i}{n} \qquad \text{(where } n = 4\text{)}$$

The $IPC_{WALS(4)}$ is strictly quantitative. It takes into account only core features whose values are highly documented. This index does not consider dependent features and values (for example, consonant-vowel ratio) and bypasses the paucity of data in certain phonological sub-areas (e.g. prosody-related features). Nevertheless, the $IPC_{WALS(4)}$ relies heavily on the experts' judgement: the incorrect description of certain features (e.g. failure to recognise vowel harmony or tonal systems) will irremediably distort the reading of the complexity measure.

5.2. Representativeness

In the previous section I have described three possible indexes. Each of these three metric tools presents advantages and disadvantages. In order to operate a choice we are

forced to look for a master criterion, which I claim to be representativeness. Even if we could assign a value to all the 20 features listed in the WALS, the obtained phonological picture is deemed to be a coarse-grain approximation. Languages can be categorised or grouped together according to the values assigned to certain features, indeed the main function of these features is dividing languages into types and not describing areas of grammar in detail. Some features are more representative than others: we could say, for example, that knowing the size of a consonant inventory is more relevant than assessing the presence or absence of velar nasals. Hence, we could tie the notion of representativeness to quantifiable inventories (consonants and vowels) and structures (syllables and tones), leaving asides binary features (velars nasals, front rounded vowels, etc.). In other words, representative features are independent, i.e. they express inventories and structures and not parts of them. Therefore, among the three indexes proposed here, the $IPC_{WALS(4)}$ (which, at this point, could be renamed IPC_4) is the most representative in that it considers only the most important (i.e. basic and quantifiable) features.

6. Phonological complexity in West African languages: some observations

In this section I will address some issues of methodological and theoretical order arising from the application of the ICP. The first observation illustrates a general principle, labelled here "Complexity Equilibrium", while the others relate to the specific case of West African languages.

6.1. The Complexity Equilibrium

The typological data available in the WALS tell us that no language displays maximum or minimum degrees of complexity.

Consonants	Vowels	Syllables	Tones	*n* of languages
large (34+)	large (7-14)	complex	complex	0
mod. small. (15-18)	large (7-14)	complex	complex	0
small (6-14)	large (7-14)	complex	complex	0
small (6-14)	small (2-4)	simple	simple	0

Table 15 – Maximum/minimum complexity (across languages present in WALS for which data are available for the 4 phonological factors)

While it is perfectly normal to find systems (or factors) with either a very small or a very large number of elements, when it comes to the configuration of systems (i.e. all the systems relevant, say, to phonology) the high complexity of one factor must be balanced by the low complexity of another factor. In this sense, a macro-system (e.g. phonology, morphology) can be neither 'overloaded' nor 'under loaded'.

6.2. Compensation effect

Maddieson, in his statistical analysis on the relationship between syllable structures, segment inventories and tone contrasts observes that no significant 'compensation effect' can be found between an increase of complexity in one sector of phonology and a decrease of complexity in another sector, the only exception being the relationship between tonal systems and syllabic structures (Maddieson 2007: 93ff.). He also states that "increasing complexity of tone system is positively associated with increasing size of both consonant and vowel inventories" (Maddieson 2007: 102).

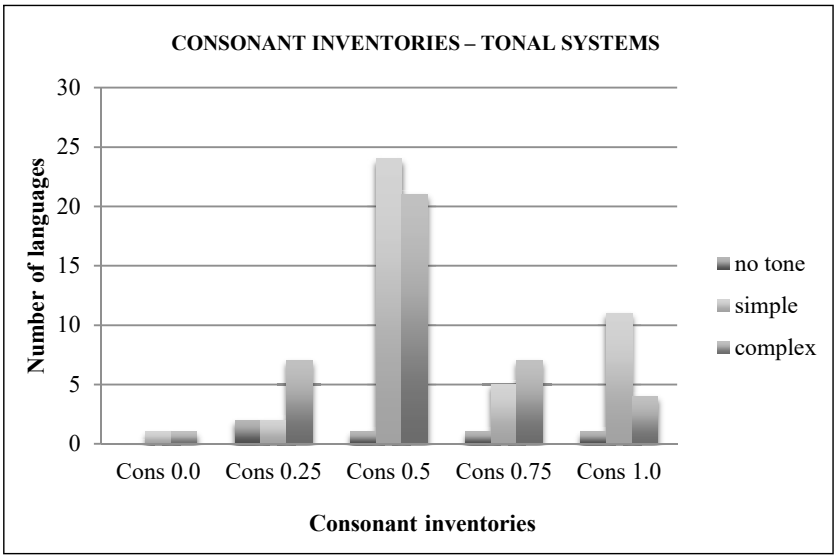

Figure 1 – Consonant inventories and tonal systems

In this sense, West African languages seem to diverge from the general pattern identified by Maddieson. Out of 98 West African languages (WALS plus other sources), 46 languages present average consonant inventories. Of these, 45 have either a simple tonal system (two-way contrast) or a complex tonal system (number of tones ≥ 3). The total number of languages with a large consonant inventory is 16, 11 of which display a simple tonal system and 4 a complex tonal system.

6.3. Availability of data

The tables displayed in the Appendix consider exclusively the languages for which the WALS reports the four core features, disregarding all those languages without a description of consonant inventory, vowel inventory, syllable structure, and tonal system. The West African languages listed in the WALS that satisfy this criterion amount to 56, which is a relatively small number if compared to the total number of West African languages represented in the *Atlas*. Among the languages for which the WALS does not provide any phonological data there is a high number of under-described and undescribed languages, but also vehicular languages such as Kanuri or Pulaar. Although the $IPC_{WALS(4)}$ requires

a limited number of features, these are not always readily available in the WALS. Fortunately, core features can be easily assessed by consulting (when existing!) grammatical sketches, articles, fieldwork reports, and conference papers.[3] One of the major features of the WALS is that all languages are geolocalised (although not necessarily in a precise manner), hence it is possible to sort them by feature or combination of features and have them represented in a map according to the values of the selected features. Considering both the WALS and a certain number of external sources, it seems that there is a general lack of data concerning the prosodic aspects of languages. Table A in the Appendix shows how the features "fixed stress locations" (14A), "weight sensitive stress" (15A), "weight factors in weight-sensitive stress systems" (16A), and "rhythm types" (17A) are rarely described. Other features that are left undescribed in the WALS do not entail paucity of data: the absence of description for certain features has more to do with the way the data has been collected and/or with the timing of its insertion in the database than with the nature of the data or the source itself. If we have at our disposal the values for features such as consonant and vowel inventories, then it is reasonable to think that the lack of values regarding the presence or the absence of velar nasals (9A), vowel nasalisation (10A), and nasal vowels (10B) is merely accidental (in other words, this is a case where the lack of information can be overcome by accessing the primary source).

6.4. West Africa as a phonological area

The purpose of this paragraph is to summarise the areal aspects from the perspective of phonological complexity as resulting from the application of the IPC (see Appendix). Table A shows the results of the $IPC_{WALS(20)}$ across the 56 West African languages listed in the WALS for which at least the core features are valued (that is, languages with no indication of any of the core features have been ignored). The languages score from a maximum of 0.700 (Doyayo) to a minimum of 0.269 (Ifik). Using different IPC, languages not only score differently but are also grouped differently (Appendix, table B). The application of the $IPC_{WALS(20)}$ results in 31 positions: languages scoring 0.700 (1 member: Doyayo), languages scoring 0.615 (4 members: Angas, Gwari, Kpelle, Tera), and so on. The $IPC_{WALS(13)}$ results in 20 positions, while the $IPC_{WALS(4)}$ results in just 12. For obvious reasons there is an observable general correspondence between the $IPC_{WALS(20)}$ and the $IPC_{WALS(13)}$. Nevertheless, if we take into account the $IPC_{WALS(20)}$ and the $IPC_{WALS(4)}$ we will observe, at least for certain languages, a certain degree of idiosyncrasy. Fe'fe (Volta-Congo), for example, scores 0.346 (29th/31) with the $IPC_{WALS(20)}$ (max. value 0.700, min. value 0.269) and 0.688 (7th/12) with the $IPC_{WALS(4)}$ (max. value 0.938, min. value 0.313). Is the phonology of Fe'fe non-complex, as the $IPC_{WALS(20)}$ seems to suggest, or is this language relatively complex, as the calculation with $IPC_{WALS(4)}$ implies? Again, the entire issue goes down to the representativeness of features. The reason for such a low score with the $IPC_{WALS(20)}$ resides in the fact that the values of a few non-core features lower the global score in a significant way. This does not mean that phonological complexity cannot be properly measured, bur rather that more than one index is possible and that different indexes are based on different phonological features (even if, as argued before, an index based on core features appears to be the best

[3] Bibliography of data-related literature not present in the WALS is listed at the end of the article.

candidate in order to attain a global, objective quantification of complexity). West African languages are consistent in their syllabic structures, which generally are moderately simple, while showing a less homogeneous pattern in consonant and vowel inventories and in tonal systems.

7. Conclusions

In this paper I have argued for the existence of a tool – in *Index of Phonological Complexity* – capable of measuring the phonological complexity of any given language. I have used the index, based on the phonological features listed in the WALS, to calculate the phonological complexity in a certain number of West African languages. The main purpose of the IPC is to calculate phonological complexity cross-linguistically by taking into account the data provided by the *highest number of languages possible.* As we have seen, the WALS, while offering a solid typological framework to build the metrics of the IPC, does not include all the data concerning phonologies of individual languages that are available in the literature. Hence, to shift from a macro-scale scenario to a micro-scale scenario, and then to a detailed representation of the latter, more data (both from existing material and from the field) are needed. A general grasp on complexity is, at least within certain limits, quite intuitive in the context of West Africa, but specific – regional and areal – dynamics require an up-to-date and detailed representation. IPC values, in order to say something meaningful, need to be statistically sound and anchored to phenomena and conditions that can play a role in the increase or decrease of complexity: geographical setting, language contact dynamics, and population size, just to name a few. In this sense, West Africa is a particularly challenging area. The lack basic phonological descriptions for a high number of languages hinders a meaningful representation of complexity, especially as far as extra-linguistic factors are concerned.

References

Anderson, Gregory D.S. 2013. "The Velar Nasal". In: M.S. Dryer, M. Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/9, Accessed on 2018-08-24.

Baechler, Raffaela & Guido Siler. 2016. *Complexity, Isolation, and Variation*. Berlin & Boston: de Gruyter.

Bentz, Christian, Tatyana Ruzsics, Alexander Koplenig, & Tanja Samardžić. 2016. "A Comparison Between Morphological Complexity Measures: Typological Data vs. Language Corpora". In: Dominique Brunato, Felice Dell'Orletta, Giulia Venturi, Thomas François, and Philippe Blache (eds.). *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity*. Osaka: The COLING 2016 Organizing Committee, 143-153.

Clements, Nickerson G. & Annie Rialland. 2007. "Africa as a phonological area". In: Bernd Heine, Derek Nurse (eds.), *A Linguistic Geography of Africa*. Cambridge: Cambridge University Press, 36-85.

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/3, Accessed on 2018-09-18.

Goedemans, Rob & Harry van der Hulst. 2013a. "Fixed Stress Locations". In: Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/14, Accessed on 2018-08-24.

Goedemans, Rob and Harry van der Hulst. 2013b. “Weight-sensitive stress”. In: Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/15, Accessed on 2018-08-24.

Goedemans, Rob and Harry van der Hulst. 2013c. “Weight factors in weight-sensitive stress systems’. In: Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/16, Accessed on 2018-08-24.

Goedemans, Rob and Harry van der Hulst. 2013d. “Rhythm types. Fixed Stress Locations”. In: Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/17, Accessed on 2018-08-24.

Hajek, John. 2013. “Velar nasalization”. In: Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/10, Accessed on 2018-08-24.

Harmon, David & Jonathan Loh. 2010. “The Index of Linguistic Diversity: A New Quantitative Measure of Trends in the Status of the World’s Languages”. *Language Documentation & Conservation* 4, 97-151.

Herder, Johannes Gottfried. 1772. *Abhandlung über den Ursprung der Sprache: welche den von der Königl. Academie der Wissenschaften für das Jahr 1770 gesezten Preis erhalten hat*. Berlin: Voss.

Hockett, Charles Francis. 1958. *A Course in Modern Linguistics*. New York: Macmillan.

von Humboldt, Wilhelm. 1836. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwickelung des Menschengeschlechts*. Berlin: Ferdinand Dümmler.

Maddieson, Ian. 2005. “Issues of phonological complexity: Statistical analysis of the relationship between syllable structures, segment inventories and tone contrasts”. In: *UC Berkeley Phonology Lab Annual Report*. Berkeley: University of California, 269-268

Maddieson, Ian. 2006. “Correlating phonological complexity: Data and validation”. *Linguistic Typology* 10(1), 106-123

Maddieson, Ian. 2007. “Issues of Phonological Complexity Statistical Analysis of the Relationship between Syllable Structures, Segment Inventories, and Tone Contrasts”. In: Maria-Josep Solé, Patrice Speeter Beddor, and Manjari Ohala (eds.). *Experimental Approaches to Phonology*. Oxford: Oxford University Press, 93-103.

Maddieson, Ian. 2011. “Phonological complexity in linguistic patterning”. In: *Proceedings of the 17th International Congress of Phonetic Sciences (Hong Kong, August 17-21)*: 28-34.

Maddieson, Ian. 2013a. “Consonant Inventories”. In: Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/1, Accessed on 2018-08-24.

Maddieson, Ian. 2013b. “Vowel quality inventories”. In: Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/2, Accessed on 2018-08-24.

Maddieson, Ian. 2013c. “Consonant-vowel ratio”. In: Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/3, Accessed on 2018-08-24.

Maddieson, Ian. 2013d. “Voicing in plosives and fricatives”. In: Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/4, Accessed on 2018-08-24.

Maddieson, Ian. 2013e. “Voicing in gaps and plosive systems”. In: Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/5, Accessed on 2018-08-24.

Maddieson, Ian. 2013f. “Uvular consonants”. In: Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/6, Accessed on 2018-08-24.

Maddieson, Ian. 2013g. “Glottalized consonants”. In: Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/7, Accessed on 2018-08-24.

Maddieson, Ian. 2013h. "Lateral consonants". In: Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/8, Accessed on 2018-08-24.

Maddieson, Ian. 2013i. "Front rounded vowels". In: Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/11, Accessed on 2018-08-24.

Maddieson, Ian. 2013j. "Syllable structure". In: Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/12, Accessed on 2018-08-24.

Maddieson, Ian. 2013k. "Tone". In: Matthew S. Dryer & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/13, Accessed on 2018-08-24.

Miestamo, Matti, Kaius Sinnemäki & Fred Karlsson. 2008. *Language Complexity. Typology, contact, change*. Amsterdam & Philadelphia: John Benjamins.

Pallotti, Gabriele. 2015. "A simple view of linguistic complexity". *Second Language Research* 31(1), 117-134.

Parkvall, Mikael. 2008. "The simplicity of creoles in a cross-linguistic perspective". In: Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.). *Language Complexity. Typology, contact, change*. Amsterdam & Philadelphia: John Benjamins, 265-286.

Sampson, Geoffrey, David Gil & Peter Trudgill. 2009. *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press.

Bibliography of data-related literature not present in the WALS

Alio, Khalil & Herrmann Jungraithmayr. 1989. *Lexique bidiya*. Frankfurt am Main: Vittorio Klostermann.

Andreas, Heike, Rudolf Leger & Ulrike Zoch. 2009. "The Nyam language: First steps toward a grammatical description". In: E. Rothmaler (ed.). *Topics in Chadic Linguistics V. Comparative and Descriptive Studies*. Köln: Rüdger Köppe Verlag, 23-36.

Baerman, Matthew, Dunstan Brown & Greville G. Corbett. 2009. "*Morphological Complexity: a typological perspective"*. In: *European Science Foundation Exploratory Workshop on Words in Action*, 2009-10-12. Istituto di Linguistica Computazionale.

Baerman, Matthew, Dunstan Brown & Greville G. Corbett. 2017. *Morphological Complexity*. Cambridge: Cambridge University Press.

Barreteau, Daniel. 1988. *Description du mofu-gudur. Langue de la famille tchadique parlée au Cameroun*. Paris: ORSTOM.

Barreteau, Daniel & Yves Le Bleis. 1990. *Lexique mafa. Langue de la famille tchadique parlée au Cameroun*. Paris: Geuthner-ORSTOM.

Batic, Gian Claudio. 2014. *A Grammatical Sketch of Bure*. Köln: Rüdiger Köppe Verlag.

Blench, Roger. 1998, 2004. *Dagbani-English Dictionary* (ms). Available online at http://www.rogerblench.info/Language/Niger-Congo/Gur/Dagbani%20dictionary%20CD.pdf, Accessed on 29.07.2018.

Blench, Roger. 2005. *The Kayigang [=Keiga, Deiga] Language of the Nuba Hills, Sudan.* Kay Williamson Educational Foundation (UK). Available online at http://www.rogerblench.info/Language/Nilo-Saharan/Kadu/Kayigang%20wordlist.pdf, Accessed on 15.12.2017.

Blench, Roger. 2007. The Dyarim language of central Nigeria and its affinities. In: Henry Tourneux (ed.). *Topics in Chadic Linguistics IV. Comparative and Descriptive Studies*. Köln: Rüdger Köppe Verlag, 41-60.

Blench, Roger. 2009. *Bura phonology and some suggestions concerning the orthography*, Kay Williamson Educational Foundation (UK). Available online at http://www.rogerblench.info/Language/Afroasiatic/Chadic/Central/Bura/Bura%20phonology.pdf, Accessed on 16.12.2017.

Blench, Roger. 2013a. *The Nupoid Languages of West-Central Nigeria: Overview and Comparative Wordlist*, Kay Williamson Educational Foundation (UK). Available online at http://www.rogerblench.info/Language/Niger-Congo/VN/Nupoid/ Nupoid%20Overview%202013.pdf, Accessed on 15.12.2017.

Blench, Roger. 2013b. *Ywom wordlist*, Kay Williamson Educational Foundation (UK). Available online at http://www.rogerblench.info/Language/Afroasiatic/Chadic/West/Ywom/Ywom%20wordlist.pdf, Accessed on 16.12.2017.

Blench, Roger & Antony Damsai. 2009. “An introduction to Kirya-Konzəl”. In: Eva Rothmaler (ed.). *Topics in Chadic Linguistics V. Comparative and Descriptive Studies*. Köln: Rüdger Köppe Verlag, 75-85.
Brindle, Jonathan (2017). *A Dictionary and Grammatical Outline of Chakali*. Berlin: Language Science Press.
Frajzyngier, Zygmunt. 1989. *A Grammar of Pero*. Berlin: Dietrich Verlag Berlin.
Friesen, Dianne. 2017. *A Grammar of Moloko*. Berlin: Language Science Press.
Guest, Elizabeth. 1997. *Heiban Grammar*, ms.
Jungraithmayr, Hermann. 1991. A Dictionary of the Tangale Language. Berlin: Dietrich Reimer Verlag.
Kaki, Aaid Komi, Hassan Ali & Ahmad Ibrahim. 1998. *Phonology write-up of Hadara (Lote)*, ms.
Kastenholz, Raimund. 2017. *La Langue Bolgo du Guéra (Tchad): Notes de recherche et matériel lexical*. [Arbeitspapiere des Instituts für Ethnologie und Afrikastudien der Johannes Gutenberg-Universität Mainz, 172.] Mainz: Institut für Ethnologie und Afrikastudien, Johannes Gutenberg-Universität.
Kidda, Mairo Elinor. 1985. *Tangale Phonology: A Descriptive Analysis (Nigeria)*. Doctoral dissertation, University of Illinois at Urbana-Champaign.
Leger, Rudolf. 2014. “Some observations on typological features in languages of the Bole-Tangale group”. In: Anne Storch, Johannes Harnischfeger, & Rudolf Leger (eds.). *Fading delimitations. Multilingual settlements in a convergence area – Case studies from Nigeria*. Köln: Rüdiger Köppe Verlag, 229-262.
Morton, Deborah. 2012. “[ATR] Harmony in an Eleven Vowel Language: The Case of Anii”. In: Michael R. Marlo, Nikki B. Adams, Christopher R. Green, Michelle Morrison, & Tristan M. Purvis (eds.). *Selected Proceedings of the 42nd Annual Conference on African Linguistics*. Somerville, MA: Cascadilla Proceedings Project, 70-78.
Roberts, James S. 2009. “Palatalization and labialization in Mawa (eastern Chadic)”. In: Eva Rothmaler (ed.). *Topics in Chadic Linguistics V. Comparative and Descriptive Studies*. Köln: Rüdger Köppe Verlag, 129-140.
Sampson, Geoffrey. 2009. “A linguistic axiom challenged”. In: Geoffrey Sampson, David Gil & Peter Trudgill (eds.). *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 1-18.
Schadeberg, Thilo C. 1985. *A Small Sketch of Ewe* [Afrikanistische Arbeitspapiere: Sondernummer, 1985]. Köln: Institut für Afrikanistik.
Schaefer, Ronald P. & Francis O. Egbokhare. 2007. *A Dictionary of Emai. An Edoid Language of Nigeria*. Köln: Rüdiger Köppe Verlag.
Schiattarella, Valentina. 2017. *Berber Texts from Siwa (Egypt)*. Köln: Rüdiger Köppe Verlag.
Zima, Petr. 1994. *Lexique dendi (Songhay) avec un index français-dendi*. Köln: Rüdiger Köppe Verlag.

Appendix

LANGUAGE	ISO	FAMILY	1A	2A	3A	4A	5A	6A	7A	8A	9A	10A	10B	11A	12A	13A	14A	15A	16A	17A	18A	19A	ICP	n.
Doyayo	dow	Benue-Congo	0.50	1.0	0.50	1.0	1.0	0.0	1.0	0.5	–	1.0	1.00	0.0	1.0	1.0	–	–	–	–	0.0	1.0	**0.700**	15
Angas	anc	Chadic	0.75	1.0	0.75	1.0	1.0	0.0	1.0	0.5	–	–	–	0.0	1.0	1.0	–	–	–	–	0.0	0.0	**0.615**	13
Gwari	gbr	Benue-Congo	0.50	0.5	0.50	1.0	1.0	0.0	1.0	0.5	–	–	–	0.0	1.0	1.0	–	–	–	–	0.0	1.0	**0.615**	13
Kpelle	xpe	Mande	0.50	1.0	0.50	1.0	1.0	0.0	1.0	0.5	–	–	–	0.0	0.5	1.0	–	–	–	–	0.0	1.0	**0.615**	13
Tera	ttr	Chadic	1.00	0.5	1.00	1.0	1.0	0.0	1.0	1.0	–	–	–	0.0	0.5	1.0	–	–	–	–	0.0	0.0	**0.615**	13
Kohumono	bcs	Volta-Congo	0.75	1.0	0.50	1.0	1.0	0.0	1.0	0.0	–	–	–	0.0		1.0	–	–	–	–	0.0	1.0	**0.604**	13
Aizi	ahp	Volta-Congo	0.50	1.0	0.25	1.0	1.0	0.0	1.0	0.5	–	–	–	0.0	1.0		–	–	–	–	0.0	1.0	**0.604**	12
Tarok	yer	Volta-Congo	0.75	0.5	0.50	1.0	1.0	0.0	1.0	0.5	–	–	–	0.0	0.5	1.0	–	–	–	–	0.0	1.0	**0.596**	13
Bété	bev	Volta-Congo	0.50	1.0	0.00	1.0	1.0	0.0	1.0	0.5	–	–	–	0.0	0.5	1.0	–	–	–	–	0.0	1.0	**0.577**	13
Dan	dnj	Mande	0.50	1.0	0.00	1.0	1.0	0.0	1.0	0.5	–	–	–	0.0	0.5	1.0	–	–	–	–	0.0	1.0	**0.577**	13
Mumuye	mzm	Volta-Congo	0.50	0.5	0.50	1.0	1.0	0.0	1.0	0.5	–	–	–	0.0	0.5	1.0	–	–	–	–	0.0	1.0	**0.577**	13
Mambila	mcu	Benue-Congo	0.50	1.0	0.25	0.5	1.0	0.0	1.0	–	–	–	–	0.0	0.5	1.0	–	–	–	–	0.0	1.0	**0.563**	13
Gã	gaa	Volta-Congo	0.75	1.0	0.50	1.0	1.0	0.0	0.0	0.5	–	–	–	0.0	0.5	1.0	–	–	–	–	0.0	1.0	**0.558**	13
Ogbia	ogb	Volta-Congo	0.50	1.0	0.25	1.0	1.0	0.0	1.0	0.5	–	–	–	0.0	0.5	0.5	–	–	–	–	0.0	1.0	**0.558**	13
Konyagi	cou	Atlantic-Congo	1.00	–	–	1.0	1.0	0.0	1.0	0.5	–	–	–	–	–	0.5	–	–	–	–	0.0	0.0	**0.556**	9
Ewe	ewe	Volta-Congo	0.75	1.0	0.50	1.0	1.0	0.0	0.0	0.5	1.0	–	–	0.0	0.5	0.5	–	–	–	–	0.0	1.0	**0.554**	14
Bagirmi	bmi	Central Sudanic	0.75	1.0	0.50	0.5	1.0	0.0	1.0	0.5	1.0	1.0	–	0.0	0.5	1.0	1.0	0.0	0.0	–	0.0	0.0	**0.542**	18
Birom	bom	Volta-Congo	0.50	1.0	0.50	1.0	1.0	0.0	0.0	0.5	–	–	–	0.0	0.5	1.0	–	–	–	–	0.0	1.0	**0.538**	13
Ngizim	ngi	Chadic	1.00	0.5	1.00	1.0	1.0	0.0	1.0	1.0	0.0	–	–	0.0	0.5	0.5	–	–	–	–	0.0	0.0	**0.536**	14
Ewondo	ewo	Volta-Congo	0.75	1.0	0.50	1.0	0.5	0.0	0.0	0.5	–	–	–	0.0	0.5	1.0	–	–	–	–	0.0	1.0	**0.519**	13
Ndut	ndv	Atlantic-Congo	0.75	1.0	0.50	0.5	1.0	0.0	1.0	0.5	–	–	–	0.0	0.5	1.0	–	–	–	–	0.0	0.0	**0.519**	13
Kpan	kpk	Benue-Congo	0.50	0.5	0.75	1.0	1.0	0.0	0.0	0.0	–	–	–	0.0	1.0	1.0	–	–	–	–	0.0	1.0	**0.519**	13
Tampulma	tpm	Volta-Congo	0.50	1.0	0.25	1.0	1.0	0.0	0.0	0.5	–	–	–	0.0	1.0	0.5	–	–	–	–	0.0	1.0	**0.519**	13

LANGUAGE	ISO	FAMILY	1A	2A	3A	4A	5A	6A	7A	8A	9A	10A	10B	11A	12A	13A	14A	15A	16A	17A	18A	19A	ICP	n.
Isoko	iso	Benue-Congo	0.75	1.0	0.50	1.0	1.0	0.0	0.0	0.5	–	–	–	0.0	0.5	0.5	–	–	–	–	0.0	1.0	**0.519**	13
Kotoko	ktk	Chadic	0.75	1.0	0.50	0.5	1.0	0.0	1.0	0.5	–	–	0.00	0.0	1.0	0.5	–	–	–	–	0.0	1.0	**0.517**	15
Igbo	ibo	Benue-Congo	0.75	1.0	0.50	1.0	1.0	0.0	0.0	0.5	–	–	–	0.0	0.5	0.5	–	–	–	–	0.0	1.0	**0.517**	15
Grebo	grj	Volta-Congo	0.50	1.0	0.25	0.5	1.0	0.0	0.0	0.5	–	–	–	0.0	0.0	1.0	0.0	1.0	1.0	–	0.0	1.0	**0.513**	19
Bariba	bba	Volta-Congo	0.25	1.0	0.25	1.0	1.0	0.0	0.0	0.5	–	–	–	0.0	–	1.0	–	–	–	–	0.0	1.0	**0.500**	13
Alladian	ald	Volta-Congo	0.50	1.0	0.50	1.0	1.0	0.0	0.0	0.5	–	–	–	0.0	0.5	0.5	–	–	–	–	0.0	1.0	**0.500**	13
Bobo Madaré	bbo	Mande	0.50	1.0	0.50	1.0	1.0	0.0	0.0	0.5	–	–	–	0.0	0.0	1.0	–	–	–	–	0.0	1.0	**0.500**	13
Dangaléat	daa	Chadic	0.50	1.0	0.50	1.0	1.0	0.0	1.0	0.5	–	–	–	0.0	0.5	0.5	–	–	–	–	0.0	0.0	**0.500**	13
Fyem	pym	Volta-Congo	1.00	0.5	1.00	0.5	1.0	0.0	1.0	0.5	–	–	–	0.0	0.5	0.5	–	–	–	–	0.0	0.0	**0.500**	13
Amo	amo	Benue-Congo	0.50	1.0	0.25	1.0	1.0	0.0	0.0	0.5	–	–	–	0.0	0.5	0.5	–	–	–	–	0.0	1.0	**0.481**	13
Margi	mrt	Chadic	0.75	0.0	1.00	1.0	1.0	0.0	1.0	0.5	–	–	–	0.0	0.5	0.5	–	–	–	–	0.0	0.0	**0.481**	13
Ejagham	etu	Benue-Congo	0.50	1.0	0.50	0.5	1.0	0.0	0.0	0.0	–	–	–	0.0	0.5	1.0	–	–	–	–	0.0	1.0	**0.462**	13
Kanakuru	kna	Chadic	0.50	0.5	0.50	0.5	1.0	0.0	1.0	1.0	–	–	–	0.0	0.5	0.5	–	–	–	–	0.0	0.0	**0.462**	13
Lelemi	lef	Volta-Congo	0.50	1.0	0.50	1.0	1.0	0.0	0.0	0.5	–	–	–	0.0	0.0	0.5	–	–	–	–	0.0	1.0	**0.462**	13
Temne	tem	Atlantic-Congo	0.25	1.0	0.25	0.5	0.5	0.0	0.0	0.5	–	1.0	0.75	0.0	1.0	0.0	–	–	–	–	0.0	1.0	**0.450**	15
Toro So	dts	Dogon	0.25	1.0	0.25	0.5	1.0	0.0	0.0	0.5	–	1.0	0.75	0.0	0.0	0.5	–	–	–	–	0.0	1.0	**0.450**	15
Aghem	agq	Benue-Congo	0.50	1.0	0.25	1.0	0.5	0.0	0.0	0.5	–	–	–	0.0	0.5	0.5	–	–	–	–	0.0	1.0	**0.442**	13
Ijo	ijc	Ijoid	0.50	1.0	0.25	1.0	1.0	0.0	0.0	0.5	–	–	–	0.0	0.0	0.5	–	–	–	–	0.0	1.0	**0.442**	13
Supyire	spp	Volta-Congo	0.50	1.0	0.50	1.0	0.5	0.0	0.0	0.5	1.0	–	–	0.0	0.5	1.0	1.0	0.0	0.0	–	0.0	0.0	**0.441**	17
Diola-Fogny	dyo	Atlantic-Congo	0.50	1.0	0.00	0.5	1.0	0.0	0.0	0.5	1.0	–	–	0.0	1.0	0.0	1.0	1.0	0.0	–	0.0	0.0	**0.441**	17
Kera	ker	Chadic	0.50	0.5	0.50	1.0	1.0	0.0	1.0	0.5		–	–	0.0	0.5	1.0	0.0	0.0	1.0	0.0	0.0	0.0	**0.441**	17
Bambara	bam	Mande	0.50	1.0	0.50	1.0	1.0	0.0	0.0	0.5	1.0	–	–	0.0	0.0	0.5	–	–	–	–	0.0	0.0	**0.429**	14
Kisi	kss	Atlantic-Congo	0.25	1.0	0.25	0.0	1.0	0.0	1.0	0.5	–	–	–	0.0	0.5	1.0	–	–	–	–	0.0	0.0	**0.423**	13
Dagbani	dag	Volta-Congo	0.50	0.5	0.75	1.0	1.0	0.0	0.0	0.5	–	0.0	0.00	0.0	0.5	0.5	–	–	–	–	0.0	1.0	**0.417**	15

LANGUAGE	ISO	FAMILY	1A	2A	3A	4A	5A	6A	7A	8A	9A	10A	10B	11A	12A	13A	14A	15A	16A	17A	18A	19A	ICP	n.
Noni	nhu	Volta-Congo	0.50	1.0	0.25	0.5	0.5	0.0	0.0	0.5	–	–	–	0.0	1.0	1.0	–	–	–	–	0.0	0.0	**0.404**	13
Kanuri	knc	Saharn	0.50	1.0	0.50	1.0	0.5	0.0	0.0	0.5	1.0	0.0	–	0.0	0.5	0.5	–	–	–	–	0.0	0.0	**0.400**	15
Hausa	hau	Chadic	0.75	0.5	0.75	1.0	0.5	0.0	1.0	0.5	0.0	0.0	–	0.0	0.5	0.5	–	–	–	–	0.0	0.0	**0.400**	15
Yoruba	yor	Benue-Congo	0.25	1.0	0.25	0.5	0.5	0.0	0.0	0.5	0.0	1.0	–	0.0	0.0	1.0	–	–	–	–	0.0	1.0	**0.400**	15
Klao	klu	Volta-Congo	0.00	1.0	0.00	0.5	0.5	0.0	0.0	0.0	–	–	–	0.0	0.0	1.0	–	–	–	–	1.0	1.0	**0.385**	13
Wolof	wlf	Atlantic-Congo	0.75	1.0	0.50	0.5	1.0	0.0	0.0	0.5	–	0.0	–	0.0	1.0	0.0	–	–	–	–	0.0	0.0	**0.375**	14
Fe'fe'	fmp	Volta-Congo	0.25	1.0	0.25	1.0	0.5	0.0	0.0	0.0	–	–	–	0.0	0.5	1.0	–	–	–	–	0.0	0.0	**0.346**	13
Akan	aka	Volta-Congo	0.50	1.0	0.25	0.5	1.0	0.0	0.0	0.0	–	–	–	0.0	0.5	0.5	–	–	–	–	0.0	0.0	**0.327**	13
Bisa	bib	Mande	0.25	0.5	0.50	1.0	1.0	0.0	0.0	0.5	–	–	–	0.0	0.5	0.0	–	–	–	–	0.0	0.0	**0.327**	13
Efik	efi	Benue-Congo	0.00	1.0	0.00	0.5	0.0	0.0	0.0	0.0	–	–	–	0.0	0.5	0.5	–	–	–	–	0.0	1.0	**0.269**	13

Table A – $IPC_{WALS(20)}$ for 56 WA languages listed in the WALS

#	Language	ISO	$IPC_{WALS(20)}$
1	Doyayo	dow	**0.700**
2	Angas	anc	**0.615**
2	Gwari	gbr	**0.615**
2	Kpelle	xpe	**0.615**
2	Tera	ttr	**0.615**
3	Kohumono	bcs	**0.604**
3	Aizi	ahp	**0.604**
4	Tarok	yer	**0.596**
5	Bété	bev	**0.577**
5	Dan	dnj	**0.577**
5	Mumuye	mzm	**0.577**

#	Language	ISO	$IPC_{WALS(13)}$
1	Doyayo	dow	**0.654**
2	Angas	anc	**0.615**
2	Gwari	gbr	**0.615**
2	Kpelle	xpe	**0.615**
2	Tera	ttr	**0.615**
3	Kohumono	bcs	**0.604**
3	Aizi	ahp	**0.604**
4	Kotoko	ktk	**0.596**
4	Tarok	yer	**0.596**
5	Bété	bev	**0.577**
5	Dan	dnj	**0.577**

#	Language	ISO	$IPC_{WALS(4)}$
1	Angas	anc	**0.938**
2	Kohumono	bcs	**0.917**
3	Doyayo	dow	**0.875**
3	Noni	nhu	**0.875**
4	Aizi	ahp	**0.833**
5	Gã	gaa	**0.813**
5	Bagirmi	bmi	**0.813**
5	Ewondo	ewo	**0.813**
5	Ndut	ndv	**0.813**
5	Kotoko	ktk	**0.813**
6	Gwari	gbr	**0.750**

#	Language	ISO	$IPC_{WALS(20)}$
6	Mambila	mcu	**0.563**
7	Gã	gaa	**0.558**
7	Ogbia	ogb	**0.558**
8	Konyagi	cou	**0.556**
9	Ewe	ewe	**0.554**
10	Bagirmi	bmi	**0.542**
11	Birom	bom	**0.538**
12	Ngizim	ngi	**0.536**
13	Ewondo	ewo	**0.519**
13	Ndut	ndv	**0.519**
13	Kpan	kpk	**0.519**
13	Tampulma	tpm	**0.519**
13	Isoko	iso	**0.519**
14	Kotoko	ktk	**0.517**
14	Igbo	ibo	**0.517**
15	Grebo	grj	**0.513**
16	Bariba	bba	**0.500**
16	Alladian	ald	**0.500**
16	Bobo Mad.	bbo	**0.500**
16	Dangaléat	daa	**0.500**
16	Fyem	pym	**0.500**
17	Amo	amo	**0.481**
17	Margi	mrt	**0.481**
18	Ejagham	etu	**0.462**

#	Language	ISO	$IPC_{WALS(13)}$
5	Mumuye	mzm	**0.577**
5	Ngizim	ngi	**0.577**
6	Mambila	mcu	**0.563**
7	Gã	gaa	**0.558**
7	Ogbia	ogb	**0.558**
8	Konyagi	cou	**0.556**
9	Birom	bom	**0.538**
10	Bagirmi	bmi	**0.519**
10	Ewondo	ewo	**0.519**
10	Ndut	ndv	**0.519**
10	Kpan	kpk	**0.519**
10	Tampulma	tpm	**0.519**
10	Ewe	ewe	**0.519**
10	Isoko	iso	**0.519**
10	Igbo	ibo	**0.519**
11	Bariba	bba	**0.500**
11	Alladian	ald	**0.500**
11	Bobo Mad.	bbo	**0.500**
11	Dangaléat	daa	**0.500**
11	Fyem	pym	**0.500**
11	Kera	ker	**0.500**
12	Amo	amo	**0.481**
12	Dagbani	dag	**0.481**
12	Margi	mrt	**0.481**

#	Language	ISO	$IPC_{WALS(4)}$
6	Kpelle	xpe	**0.750**
6	Tera	ttr	**0.750**
6	Bété	bev	**0.750**
6	Dan	dnj	**0.750**
6	Mambila	mcu	**0.750**
6	Konyagi	cou	**0.750**
6	Birom	bom	**0.750**
6	Kpan	kpk	**0.750**
6	Tampulma	tpm	**0.750**
6	Bariba	bba	**0.750**
6	Ejagham	etu	**0.750**
6	Supyire	spp	**0.750**
7	Tarok	yer	**0.688**
7	Ewe	ewe	**0.688**
7	Isoko	iso	**0.688**
7	Igbo	ibo	**0.688**
7	Kisi	kss	**0.688**
7	Wolof	wlf	**0.688**
7	Fe'fe'	fmp	**0.688**
8	Mumuye	mzm	**0.625**
8	Ogbia	ogb	**0.625**
8	Ngizim	ngi	**0.625**
8	Grebo	grj	**0.625**
8	Alladian	ald	**0.625**

#	Language	ISO	$IPC_{WALS(20)}$
18	Kanakuru	kna	**0.462**
18	Lelemi	lef	**0.462**
19	Temne	tem	**0.450**
19	Toro So	dts	**0.450**
20	Aghem	agq	**0.442**
20	Ijo	ijc	**0.442**
21	Supyire	spp	**0.441**
21	Diola-Fogny	dyo	**0.441**
21	Kera	ker	**0.441**
22	Bambara	bam	**0.429**
23	Kisi	kss	**0.423**
24	Dagbani	dag	**0.417**
25	Noni	nhu	**0.404**
26	Kanuri	knc	**0.400**
26	Hausa	hau	**0.400**
26	Yoruba	yor	**0.400**
27	Klao	klu	**0.385**
28	Wolof	wlf	**0.375**
29	Fe'fe'	fmp	**0.346**
30	Akan	aka	**0.327**
30	Bisa	bib	**0.327**
31	Efik	efi	**0.269**

#	Language	ISO	$IPC_{WALS(13)}$
13	Ejagham	etu	**0.462**
13	Hausa	hau	**0.462**
13	Kanakuru	kna	**0.462**
13	Lelemi	lef	**0.462**
14	Grebo	grj	**0.442**
14	Aghem	agq	**0.442**
14	Ijo	ijc	**0.442**
15	Supyire	spp	**0.423**
15	Kisi	kss	**0.423**
16	Noni	nhu	**0.404**
16	Wolof	wlf	**0.404**
17	Kanuri	knc	**0.385**
17	Temne	tem	**0.385**
17	Yoruba	yor	**0.385**
17	Bambara	bam	**0.385**
17	Klao	klu	**0.385**
17	Toro So	dts	**0.385**
18	Fe'fe'	fmp	**0.346**
18	Diola-Fogny	dyo	**0.346**
19	Akan	aka	**0.327**
19	Bisa	bib	**0.327**
20	Efik	efi	**0.269**

#	Language	ISO	$IPC_{WALS(4)}$
8	Bobo Madaré	bbo	**0.625**
8	Dangaléat	daa	**0.625**
8	Fyem	pym	**0.625**
8	Amo	amo	**0.625**
8	Aghem	agq	**0.625**
8	Diola-Fogny	dyo	**0.625**
8	Kera	ker	**0.625**
8	Kanuri	knc	**0.625**
8	Akan	aka	**0.625**
9	Temne	tem	**0.563**
9	Hausa	hau	**0.563**
9	Yoruba	yor	**0.563**
10	Kanakuru	kna	**0.500**
10	Lelemi	lef	**0.500**
10	Ijo	ijc	**0.500**
10	Bambara	bam	**0.500**
10	Dagbani	dag	**0.500**
10	Klao	klu	**0.500**
10	Efik	efi	**0.500**
11	Margi	mrt	**0.438**
11	Toro So	dts	**0.438**
12	Bisa	bib	**0.313**

Table B – $IPC_{WALS(20)}$ vs $IPC_{WALS(13)}$ vs $IPC_{WALS(4)}$