

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 Tuvinian² 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, Tuaric (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.

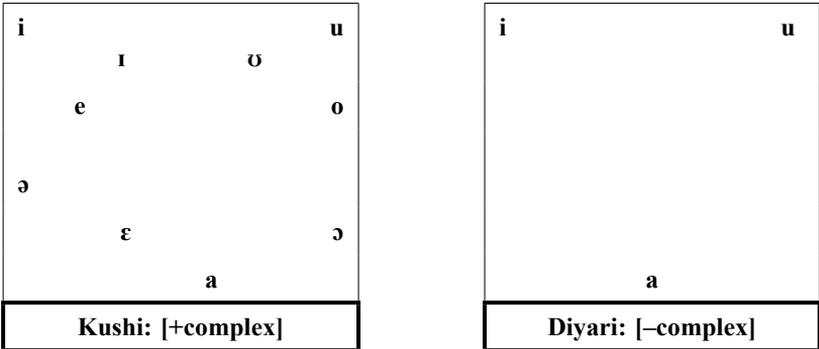


Table 1 – Vowel systems of Kushi (Afroasiatic, Chadic / Nigeria) and Diyari (Pama-Nyungan / Australia): [±complexity]

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