Exploring grammatical complexity crosslinguistically: The case of gender

Francesca Di Garbo
University of Helsinki

This paper proposes a set of principles and methodologies for the crosslinguistic investigation of grammatical complexity and applies them to the in-depth study of one grammatical domain, gender. The complexity of gender is modeled on the basis of crosslinguistically documented properties of gender systems and by taking into consideration interactions between gender and two other grammatical domains: nominal number and evaluative morphology. The study proposes a complexity metric for gender that consists of six features: "Gender values", "Assignment rules", "Number of indexation (agreement) domains", "Cumulative exponence of gender and number", "Manipulation of gender assignment triggered by number/countability", and "Manipulation of gender assignment triggered by size". The metric is tested on a sample of 84 African languages, organized in subsamples of genealogically related languages. The results of the investigation show that: (1) the gender systems of the sampled languages lean towards high complexity scores; (2) languages with purely semantic gender assignment tend to lack pervasive gender indexation; (3) languages with a high number of gender distinctions tend to exhibit pervasive gender indexation; (4) some of the uses of manipulable gender assignment are only attested in languages with a high number of gender distinctions and/or pervasive indexation. With respect to the distribution of the gender complexity scores, the results show that genealogically related languages tend to have the same or similar gender complexity scores. Languages that display exceedingly low or high gender complexity scores when compared with closely related languages exhibit distinctive sociolinguistic profiles (contact, bi- or multilingualism). The implications of these findings for the typology of gender systems and the crosslinguistic study of grammatical complexity and its distribution are discussed.

1. Introduction

Investigating the complexity of individual grammatical domains from a crosslinguistic perspective is still a novel research area within language typology. This paper focuses on the empirical study of grammatical complexity and proposes a set of principles and methodologies that can be operationalized to explore linguistic complexity crosslinguistically.[1] The paper takes inspiration from the suggestions made by Miestamo (2006b, 2008) on the typological study of grammatical complexity. According to Miestamo, complexity metrics suitable for typological purposes should not aim to assess the

---

grammatical complexity of languages in their entirety (*global complexity*), but rather focus on specific domains of grammar (e.g. functional domains) as encoded across languages, and attempt to characterize "the cross-linguistic variety in the complexity of each functional domain and the interactions between domains" (2006b) (*local complexity*).

The grammatical domain that I investigate in this paper is grammatical gender. Gender is a type of nominal classification device (in the sense of Aikhenvald 2003) that is commonly associated with high degrees of complexity, inasmuch as it presupposes inflectional morphology (agreement) and rather opaque grammaticalization paths (Corbett 1991; Dahl 2004; Nichols 1992). In this study, I attempt to model the complexity of gender by identifying a set of dimensions that characterize gender systems crosslinguistically and by taking into consideration interactions and possible asymmetries between gender and two other nominal grammatical domains, number and evaluative morphology. The paper proposes a complexity metric for gender. This metric is then tested on a sample of 84 African languages. The aim of the paper is to investigate whether crosslinguistic variation in the types of gender systems attested in the sample languages is tied to certain levels of complexity, and why this might be the case. In addition, by exploring gender complexity within and across genealogical groupings, the study aims to investigate to which extent the complexity of gender – a morphosyntactic feature that is usually conceived of as very stable in the history of language families – is conservative across related languages and under which conditions it is subject to decrease or increase. The paper is structured as follows. In section 2, I define the notion of grammatical complexity that I work with. In section 3, I introduce gender as a grammatical domain and consider possible dimensions for the assessment of gender complexity. The methodology followed in the study is illustrated in section 4: section 4.1 provides an outline of the sampling procedure; section 4.2 presents the complexity metric and section 4.3 illustrates the method used to compute complexity scores for the gender systems of the sampled languages. The results are presented in section 5 and discussed in section 6, before I provide some concluding remarks in section 7.

2. Defining grammatical complexity

The idea that all languages are equally complex is known in the literature as the *equi-complexity hypothesis* and is based on the assumption that, even though individual languages may exhibit different levels of complexity in different domains of their grammars, complexity in one domain is compensated by simplicity in another domain (*complexity trade-offs*). The equi-complexity hypothesis has long been maintained as a truism within linguistic research (for an overview, see McWhorter 2001; Kusters 2003). During the past fifteen years, however, starting from the comparative study of grammatical complexity in creole and non-creole languages by McWhorter (2001), a whole body of research (see, among others, Dahl 2004; Kusters 2003; Miestamo 2006b; Miestamo et al. 2008; Sinnemäki 2011) has suggested that the equi-complexity hypothesis is difficult to test empirically and that, when tested (e.g., by McWhorter 2001), it is actually problematic to maintain. In a nutshell, this research has shown that "there is no principled reason why all languages should be equal in their overall complexity or why complexity in one grammatical area should be compensated by simplicity in another" (Miestamo 2006b). Once we acknowledge that human languages may differ in complexity,[2] and that these differences are worth exploring for a multifaceted array of purposes (typological, sociolinguistic, historical, etc.), three major challenges follow: (1) how

---

[2] For the sake of clarity, the notion of linguistic complexity that I work with in this paper is in no way related to any type of judgment about how expressive a given language is. Showing that the grammar (or aspects of the grammar) of a language is (are) simpler than that (those) of other languages by no means implies that the former language is less efficient – from the point of view of communication – or more primitive than the latter.

to define complexity; (2) how big a scope a complexity metric should have for it to be meaningful, and (3) which principles might help to assess complexity differences in one or several domains of grammar. The three issues are discussed in section 2.1, section 2.2, and section 2.3, respectively.

2.1 Absolute and relative complexity

There exist two main approaches to the study of linguistic complexity, the *relative* and the *absolute* approach (Miestamo 2008). The relative approach (also known as *user-oriented approach*) focuses on the *costs* and *difficulties* in language learning and processing. The absolute approach (also known as *theory-oriented approach*) rather views complexity as an objective property of languages. Within the absolute approach, complexity can be assessed by measuring the number of distinctions within a system/grammatical domain, and the length of its description.

Both approaches have been used, and argued for, in typologically oriented literature on grammatical complexity. Kusters (2003), for instance, defines complexity in terms of *difficulty*. In his work on the typology of verbal inflection, Kusters examines four genealogically unrelated sets[3] of closely related languages and investigates how, within each set, languages differ in the complexity of verbal inflection and what type of sociolinguistic and sociohistorical factors may account for these differences. His definition of complexity is based on the difficulties – as documented in the psycholinguistic literature on second language acquisition – that adults incur when learning a new language. According to this definition, languages that are more "adapted" to the presence of L2 learners (*exoteric* languages, following the terminology proposed by Lupyan & Dale 2010) are less complex than languages that, throughout their history, have not been exposed, or not to the same extent, to the presence of adult learners (*esoteric* languages, based on Lupyan & Dale 2010). This definition of complexity/difficulty fits well the scope of Kusters' (2003) study, which is to investigate the effects of multilingualism, asymmetrical bilingualism and adult language contact on language structures. However, as Miestamo (2006b) rightly points out, L2 learners represent only one type of language users. In addition, adult, post-critical threshold language contact is only one type of contact scenario in the history of a speech community.[4] It follows that a definition of complexity/difficulty that is targeted to one category of language users only might not be inclusive enough if our aim is to build a more general model of linguistic complexity. Finally, given our still limited knowledge of the cognitive processes behind language learning and usage, we do not have enough evidence to model the whole range of difficulties and costs that both L1 and L2 speakers and listeners experience when using language. Thus, based on our current state of knowledge, the absolute approach allows for a more general, objective, definition of the notion of complexity. This is in turn essential for the sake of crosslinguistic comparison. In addition, the absolute approach to grammatical complexity is the one that is more easily connectable with how *complexity* is approached by other disciplines (e.g., philosophy, information theory) and thus "opens possibilities for interdisciplinary research" (Miestamo 2008: 27). Advocates of the absolute approach to the typological study of grammatical complexity are, among others, McWhorter (2001); Dahl (2004); Miestamo (2006b, 2008); Nichols (2009); Sinnemäki (2011). The absolute approach is followed in this paper. Accordingly, I use the term *complexity* to refer

---

[3] The languages of each set are taken from the following groupings: Bantu, Germanic, Quechuan, Semitic.

[4] Studies of language contact have shown that while short-term language contact between adult learners is likely to lead to simplification, long-term language contact that is characterized by pre-critical threshold multilingualism (i.e., child multilingualism) is likely to lead to complexification. For a general overview, see Trudgill (2011: chapter 2).

to *absolute complexity* and the term *difficulty* to refer to *relative complexity*.

2.2 Global vs. local complexity

One issue that has been at the center of the recent debate on grammatical complexity is how big a scope a complexity metric should have for it to be meaningful. McWhorter (2001) elaborates a complexity metric that aims to measure overall differences in the grammatical complexity of creole and non-creole languages. The metric captures phonological, morphological, syntactic and semantic patterns that involve various types of redundancy (in terms of number of overt distinctions and amount of rules) and thus qualify a language as more complex than another. Two languages are investigated in the first part of the study, the highly inflectional language Tsez (Nakh-Daghestanian) and the creole language Saramaccan. The metric individuates clear-cut complexity differences between the two languages: Sarammaccan systematically qualifies as simpler than Tsez with respect to all the parameters under investigation. In the second part of the study, the same complexity metric is used to compare Saramaccan with an non-creole analytic language, Lahu (Sino-Tibetan), based on the hypothesis that "the complexity difference between creoles and analytic languages would be less than that between them and inflected languages" (McWhorter 2001: 143). Nevertheless, the comparison reveals complexity differences between Saramaccan and Lahu that are similar to those found for Tsez and Saramaccan. These results would seem to confirm McWhorter's hypothesis whereby the grammar of creole languages is systematically simpler than that of non-creole languages. The question however remains whether a metric of this type could be effectively used to capture complexity differences (1) between a higher number of languages than those considered in McWhorter's study, and (2) based on a sampling procedure that is independent of the creole/non-creole dichotomy. Developing a metric that would satisfy these conditions and would allow us to compute the total complexity of a language in typologically meaningful ways is ultimately a massive, daunting task (see also discussion in Miestamo 2006b, 2008; Nichols 2009). In addition, even if, as suggested by Nichols (2009: 111), one would be able "to draw a representative sample of complexity in enough different grammatical domains, relatively easy to survey, to give a reliable indication of whether overall complexity does or does not vary", it would be still very hard (and probably even impossible) to establish the mutual comparability between the criteria used in the metric. In other words, it would be extremely difficult to decide whether, for instance, the number of tense distinctions, phonemes, or gender distinctions that are grammaticalized in a given language contribute in the same way to the total complexity of that language. Miestamo (2006b, 2008) refers to this as *the problem of comparability* and suggests that in view of this difficulty, the crosslinguistic study of grammatical complexity should be based on individual areas of grammar, such as *functional domains*, rather than on grammars in their entirety, and thus have a *local* rather than *global* scope. In this paper, I follow this suggestion and investigate the complexity of one grammatical domain, gender. In addition, based on Dahl (2011), I argue that in order to be maximally local, complexity metrics should be based on *ceteribus paribus comparisons*, that is on statements of the type: "Everything else being equal, X is more complex than Y".

2.3 Complexity principles

In this study, I suggest that, within an absolute and local approach to grammatical complexity (see section 2.1 and 2.2), three principles can be used as general guidelines to define the variables of a complexity metric: the *Principle of Fewer Distinctions*, the *Principle of One-Meaning–One-Form* and the *Principle of Independence*. The first two principles are well established in the literature on grammatical complexity (for an overview, see Miestamo 2008). The third principle, the Principle of Independence, was introduced by

Di Garbo (2014) to account for interactions between functional domains and complexity. In the following, I outline my definitions of the three principles:

•*The Principle of Fewer Distinctions* (proposed by Miestamo 2006a, 2008 and also known as *Principle of Economy*, see e.g., Kusters 2003): Everything else being equal, a grammatical domain with *n* distinctions is less complex than one with *n+1* distinctions.

•*The Principle of One-Meaning–One-Form* (well established in the literature on theoretical morphology and linguistic complexity, also known as the *Principle of Transparency*, see, for instance, Kusters 2003): (a) Everything else being equal, a grammatical meaning with *n* forms is less complex than one with *n+1* forms; (b) Everything else being equal, a grammatical form with *n* meanings is less complex than one with *n+1* meanings.

•*The Principle of Independence* (introduced by Di Garbo 2014):[5] Everything else being equal, a grammatical domain that is *independent* of semantic and functional properties of other domains is less complex than a grammatical domain that is *dependent* on *n* or *n+1* semantic and functional properties of other grammatical domains.

The Principle of Fewer Distinctions is concerned with the type and number of grammatical meanings that a language expresses within a given domain of grammar. For instance, other things being equal, a language with more than five genders (e.g., Swahili) is more complex in this respect than a language with three genders only (e.g., German). The Principle of One-Meaning–One-Form has to do with the type of encoding of a grammatical meaning within a given domain of grammar. The Principle of One-Meaning–One-Form can be operationalized in two ways, depending on whether we consider the mapping between form and meaning or, vice versa, the mapping between meaning and form. In addition, as suggested by Miestamo (2008: 33), the relationship between form and meaning can be investigated both at the paradigmatic and syntagmatic level. For instance, with respect to the encoding of standard negation, Italian, whose standard negator is *non*, is, other things being equal, less complex than French, which typically uses a discontinuous marker, *ne...pas*, to signal standard negation. Or, similarly, other things being equal, Turkish is simpler than German with respect to the type of exponence of case and number. In Turkish, the two grammatical meanings are encoded separately (one form for each meaning), whereas in German, number and case are encoded cumulatively (one marker for several meanings).[6] Both these violations of the Principle of One-Meaning–One-Form operate on the syntagmatic level. On the other hand, phenomena such as allomorphy and syncretism represent a violation of the Principle of One-Meaning–One-Form at the paradigmatic level. Finally, the Principle of Independence models interactions between domains and their effect on complexity. For instance, a language in which gender assignment is dependent on evaluative meanings –

---

[5] The question of how to model interactions between domains has already been approached in the literature on linguistic complexity. Dahl (2004: 46-50), for instance, uses the notion of *choice structure* to explain the selection of the value of a grammatical category (e.g., case) based on the syntactic context or the speech situation in which it occurs. This issue is also approached within the framework of Canonical Typology (see, for instance, Corbett 2012: 158).

[6] With respect to number of case distinctions, and based on the Principle of Fewer Distinctions, the case system of Turkish would, of course, rank higher in complexity than the case system of German. This example is a clear illustration of the fact that, as pointed out in section 2.2, complexity evaluations can only have a local rather than global scope.

if, e.g., masculine nouns can be shifted to the feminine gender when a diminutive meaning is encoded (as in the Berber language Kabyle) – is more complex in this respect than a language in which gender assignment cannot be manipulated for such purposes (as in the Romance language Italian).

In the remainder of this paper, the Principle of Fewer Distinctions, the Principle of One-Meaning–One-Form and the Principle of Independence will be operationalized in designing a complexity metric for grammatical gender.

3. Grammatical gender and dimensions of gender complexity

3.1 Gender as a grammatical domain

In this paper, I follow the most widely accepted definition of gender within the typological literature (Corbett 1991; Hockett 1958). Thus I define gender as a type of nominal classification strategy that must be reflected beyond nouns, via agreement patterns (Di Garbo 2014: 3). Under this definition I include both systems of the Bantu type (large number of genders) and systems of the Romance type (small number of genders). Following Croft (2001, 2003, 2013), however, I refer to agreement patterns as *indexation patterns*. Accordingly, I define the entities whose inflectional morphology signals gender (e.g., pronouns, adjectives, verbs) as *gender indexes* (or *gender indexing targets*) and the entities that trigger a given gender indexation pattern (i.e., nouns, pronouns, noun phrase referents) as *indexation triggers*. In Corbett's (1991) terminology, indexes and indexation triggers are referred to as *agreement targets* and *controllers*, respectively.[7] In the remainder of this section, I provide a short overview of the criteria used for the synchronic classification of gender systems, the debate over the origins of gender, and the function(s) of gender in discourse.

Synchronically, the gender systems of individual languages are usually classified based on: (1) the number of gender distinctions (Corbett 1991, 2013a); (2) whether gender distinctions are sex-based or non-sex-based based (Corbett 1991, 2013b); (3) the criteria according to which nouns are assigned to a given gender (Corbett 1991, 2013c).

Diachronically, gender has been observed to be one of the most stable features of grammar. Gender systems are stable with respect to two of the three criteria for stability proposed by Nichols (1992): *diachronic persistence* and *areal contingency*. Gender is one of the most conservative features in the history of language families (stability as *diachronic persistence*). For instance, Armenian is the only independent branch of the Indo-European language family that has completely lost grammatical gender. In addition, gender systems exhibit a hotbed–outlier type of distribution (stability as *areal contingency*): some areas of the world, such as Africa or Australia, are densely populated by languages with gender (*gender hotbeds*), whereas in other areas of the world (e.g., North America), the feature is absent or attested only in isolated cases (*gender outliers*).

The debate over the origins of gender is very controversial and, in many respects, still unresolved. On the one hand, it has been shown that gender systems may originate from classifier systems and/or from demonstratives (Greenberg 1978; Corbett 1991). On the other hand, among the issues that are still open for debate is, for instance, the question of whether indexation or classification comes first in the diachrony of gender within a given language or language family (Nichols 1992). The main difficulty behind the reconstruction of the diachrony of gender in many language families is that, in view of their overall stability, gender systems tend to presuppose long grammaticalization paths and their origin often precedes those stages that can be reconstructed via the historical-comparative method.

---

[7] For a theoretical discussion of the term indexation as opposed to agreement see Croft (2003, 2013).

Finally, from a functional point of view gender has been defined as a grammatical device for the management of reference in discourse, its functions being often related to reference tracking (Heath 1975; Foley & Van Valin 1984) and/or discourse redundancy (Dahl 2004). The debate over the discourse functions of gender is huge and cannot be extensively surveyed here (for an overview, see Kilarski 2013: chapter 6, as well as Contini-Morava & Kilarski 2013). For the sake of this paper, suffice it to say that scholars usually disagree on whether the complex redundancies that gender indexation introduces in discourse facilitate communication (Dahl 2004) or exist beyond communicative necessity (McWhorter 2001). Evidence from second language acquisition is often brought in support of the latter argument: contact varieties that emerge as a result of intensive post-threshold language contact and nonnative acquisition tend to systematically lack gender; similarly, adult learners usually struggle with grammatical gender when acquiring a new language.

3.2 The dimensions of gender complexity

Together with verbal inflection (Kusters 2003) and core argument marking (Sinnemäki 2011), gender figures as one of the few areas of grammar that have, so far, received some attention in the literature on linguistic complexity. Perhaps this is because grammatical gender is one of the domains of grammar that most promptly leads itself to be associated with complexity, being both theoretically and empirically relevant for the study of such notions as inflectional morphology (Nichols 1992), maturity (Dahl 2004) and redundancy in information management (McWhorter 2001).

Grammatical gender, in the form of gender indexation and overt gender distinctions on nouns, is one of the features of the complexity metric proposed by Nichols (2009). In this study, properties of gender systems are surveyed together with properties of other nominal classification devices (numeral and possessive classifiers) under the label *classification*. Within the metric proposed by Nichols, presence of gender indexation and overt marking of gender on nouns feature higher degrees of complexity.[8]

A more detailed qualitative study of the dimensions of gender complexity – viewed independently of other nominal classification devices – is Audring (2014). Audring argues that the complexity of gender systems is tied to and can be investigated by taking into considerations three main dimensions: complexity of values; complexity of assignment rules; and complexity of formal marking.

Dimension 1, *complexity of values*, is concerned with the number of genders in a language: the higher the number of genders, the more complex the gender system. Dimension 2, *complexity of assignment rules*, is concerned with the type and scope of gender assignment rules. With respect to *type* of assignment rules, the literature on the typology of gender systems (Corbett 1991, 2013c) has shown that there exists two principles according to which nouns are assigned to a gender in a given language: *semantic* and *formal*. Under semantic assignment rules, gender assignment is predicted on the basis of the meaning of nouns. Under formal assignment rules, gender assignment is predicted based on morphological rules (e.g., inflectional classes, derivational morphology) and/or phonological rules. In principle, the least complex gender system is one in which only one type of assignment rule is attested, semantic or formal. In reality, typological studies of gender (Corbett 1991, 2013c) have shown that while solely semantic gender systems are relatively common among the world's languages (e.g., among Dravidian languages), gender systems purely based on formal

---

[8] Nichols' study is an attempt to test the equi-complexity hypothesis (see section 1) on a large sample of languages and based on a selection of features ranging from phonology to the lexicon. The study finds "no significant negative correlations between different components of grammar" (Nichols 2009: 119), which suggests that it is not possible to prove that complexity in one domain of grammar is compensated by simplicity in other domains. The equi-complexity is thus not supported by the data presented in the study.

assignment rules are almost never encountered. Even in those systems that are heavily skewed towards formal mechanisms of gender assignment, there is always at least a minimal portion of the nominal lexicon (often nouns denoting humans and/or animate entities) for which gender is assigned based on clear-cut semantic criteria.[9] As for the *scope* of assignment rules, this has to do with the degree of generality of a rule, that is the gender assignment of how many nouns a given rule is able to predict. The higher the number of nouns assigned to a certain gender by a given assignment rule, the larger the scope of the assignment rule. In general, a system with large assignment rules requires a lower number of rules, leading to lower complexity. These rules usually rests upon some basic semantic notions such as sex or animacy (Audring 2014: 11).

Dimension 3, *complexity of formal marking*, is concerned with the pervasiveness of gender marking in discourse, that is, via indexation. The most straightforward implementation of this dimension of the complexity of gender is to count how many gender indexes there are in a language based on how many word classes inflect for gender (e.g., pronouns, adjectives, verbs), and independently of how these inflections are realized in discourse. The higher the number of gender indexes, the greater the complexity of a gender system. However, it is also possible to explore this dimension of gender complexity by looking at discourse frequencies, that is by measuring how often gender inflections appear in a given chunk of discourse (the higher the frequency of gender marking in discourse, the more complex the system). This aspect of the complexity of gender (which will not be explored further in this paper) can also be operationalized in the investigation of the functionality of gender indexation in language learning and processing. In this sense, a particularly promising hypothesis that is put forward in Audring's (2014) work is that, pervasive gender indexation facilitates the learning and processing of gender values and assignment rules, given that users are exposed to multiple occurrences of gender marking in a given chunk of discourse.

The gender system of English would rank low with respect to all three dimensions of complexity: it has only three genders, a few semantic assignment rules, and gender indexation is restricted to the pronominal domain.

To sum up, Audring (2014) suggests that the absolute complexity of gender systems can be explored on the basis of three macro-dimensions: number of values, assignment and indexation. This suggestion is followed in the present paper. In section 4.2, I propose one way of implementing the three dimensions into a complexity metric.

4. Methodology

4.1 Sampling procedure

This study is based on a sample of 84 gendered languages selected from the African macro-area and organized in subsets of genealogically related languages (the sample languages are listed in alphabetical order in appendix A).[10] The macro-area sampled in the study, Africa, is one of the world's gender hotbeds (Nichols 1992, 2003): all major genealogical groupings within the area display gender at least at some level of their internal taxonomies. The language classification followed in the paper is the one proposed by Glottolog (Nordhoff et al. 2013) as of September, 2015.

The sample designed for this study differs from classical sampling procedures in linguistic

---

[9] The Koman language Uduk, spoken in Ethiopia, would seem to represent an exception to this otherwise universal tendency. In Uduk, semantics seems to play no role in gender assignment (for a description of the gender system of Koman, see Killian 2015).

[10] The language sample also contains languages such as Hebrew and Maltese, which are actually spoken outside Africa. As Dryer (1989: 268) points out, all Semitic languages can be seen as part of the same large linguistic area because "their genetic relationships go in that direction".

typology. Traditionally, these procedures aim to maximize the representation of linguistic diversity by contributing one datapoint (i.e., one language) per genealogical unit.[11] In recent years, statistically implemented sampling methodologies that attempt to investigate linguistic patterns as distributed within language families have been proposed, for instance, by Maslova (2000) and Bickel (2013). The main assumption behind these methodologies is that typological distributions concerning linguistic variables reflect different historical scenarios that may favor the presence/development/maintenance or, rather, the absence/decline/loss of the variables in question. Accordingly, these studies argue that it is possible to explore "statistical biases in diachronic developments on the basis of synchronic samples" (Bickel 2013: 415). The design of the present sample is built on similar assumptions. However the study does not focus on the elaboration of stochastic models of language change based on the observation of synchronic distributions. The aim of the study is, in fact, mostly descriptive. What I am looking for is the degree of grammatical complexity that is associated with gender crosslinguistically and the extent to which this complexity is genealogically and areally uniform.

The sample consists of seventeen different genealogical units (or *lineages* following the terminology by Nichols 1992), among which two isolates (Hadza and Sandawe). Some of these units represent different subgroups of the same superordinate taxonomic level (stock)[12]. In general, language selection has been guided by the following rule of thumb: the higher the diversity (in terms of number of languages/subgroups) of a superordinate genealogical unit, the higher the number of languages/subgroups selected for that unit. Consequently, the biggest and more diverse language families are represented by a number of subsamples that tends to reflect this diversity. For instance, all major subdivisions of the Afro-Asiatic stock (except Egyptian) are represented in the sample. The subsamples created for each stock should be understood as convenience samples since (1) the number of languages per genealogical units is not established mathematically and (2) for the biggest stocks, not all subdivisions are included. The latter especially applies to the largest stock within the African macro-area, Atlantic-Congo. Some relevant genealogical units, such as Kru and, from the Volta-Congo sub-branch, Gur and Ubangi are, for instance, not included in the sample mainly due to lack of accessible resources. This impacts data analysis in that the data-set created for this study cannot be used for statistical analysis of the inferential type, that is to make predictions about preferred typological patterns in the languages of Africa and beyond. Thus, as mentioned above, the statistical analysis that will be applied to the data presented in the study is purely descriptive. Table 1 illustrates the number of genealogical units/languages per stock.

---

[11] One of the most well-known and practiced methods of language selection in typology is the one designed by Dryer (1989). Dryer uses *genera* – i.e., genealogical units with time depth comparable to that of Indo-European subfamilies such as Romance or Germanic – as the basis for language selection. For an overview of sampling procedures in linguistic typology see Bakker (2011).

[12] Nichols (1992: 25) defines a stock as the "highest level reconstructable by the standard comparative method."

Superordinate/Stock level	Genealogical units	No. of lgs
Afro-Asiatic	Berber	6
	Chadic	6
	Cushitic	13
	Semitic	7
Omotic[13]	Dizoid	1
	South Omotic	1
	Ta-Ne-Omotic	4
Atlantic Congo	Bantoid, Bantu	23
	Kwa	1
	Mel	3
	North-Central Atlantic	7
Hadza		1
Khoe-Kwadi		5
Kka		1
Nilotic	Eastern Nilotic	3
Sandawe		1
Tuu		1
Total		84

Table 1. Genealogical units in the sample

4.2 The features of the complexity metric

The complexity metric that I designed for the purpose of this study consists of six features. These can be further grouped into three main domains, which are based on the three dimensions of gender complexity proposed by Audring (2014) and discussed in section 3.2: *complexity of values*, *complexity of rules* and *complexity of formal marking*. The features of the complexity metric are presented in table 2.

---

[13] The genealogical relationships between the Omotic groups and their affiliation to Afro-Asiatic are still debated issues among specialists (for an overview, see Amha 2012). For instance, Glottolog (Nordhoff et al. 2013) classifies all the Omotic groups as independent groups outside Afro-Asiatic. In the table, I use Omotic as an areal cover term and follow the Glottolog classification for the individual subgroups.

Dimension	Feature	ID	Description
VALUES	Number of gender values	GV	Everything else being equal, a gender system with two values (gender distinctions) is less complex than a gender system with more than two values.
ASSIGNMENT RULES	Number and nature of assignment rules	AR	Everything else being equal, a gender system with one type of assignment rules – e.g., only semantic or only formal – is less complex than a gender system with two types of assignment rules – both semantic and formal.[14]
ASSIGNMENT RULES — MANIPULABLE ASSIGNMENT	Triggered by number/countability	M1	Everything else being equal, a gender system where gender assignment is only lexically given is less complex than a gender system where gender assignment is given in the lexicon + can be manipulated *depending* on the countability properties of the noun or the noun phrase.
ASSIGNMENT RULES — MANIPULABLE ASSIGNMENT	Triggered by size	M2	Everything else being equal, a gender system where gender assignment is only lexically given is less complex than a gender systems where gender assignment is given in the lexicon + can be manipulated *depending* on the size of the noun phrase referent.
FORM MARKING	Number of indexation domains	IND	Everything else being equal, a gender system that has gender indexation in one domain only (e.g. only on articles or only on pronouns) is less complex than a gender system with two or more indexation domains.
FORM MARKING	Cumulative exponence of gender and number	CUM	Everything else being equal, a marker that only signals gender is less complex than a marker that signals gender + number.

Table 2. Features of the complexity metric and their description

Features GV, AR and IND can be seen as direct implementations of Audring's (2014) three dimensions of gender complexity. Complexity with respect to GV counts as a violation of the Principle of Fewer Distinctions (the higher the number of gender distinctions, the more complex the system). Less straightforward is, on the other hand, the interpretation of AR and IND with respect to the three complexity principles outlined in 2.3. Here, I propose to view complexity with respect to AR as a violation of the Principle of Independence, and complexity with respect to IND as a violation of the Principle of One-Meaning–One-Form (both on the syntagmatic and paradigmatic level) and the Principle of Independence. On the one hand, systems of gender assignment that are dependent only on semantics or only on form are less complex than systems of gender assignment that are dependent both on semantics and form (violation of Principle of Independence). On the other hand, in a language in which many word classes inflect for gender, and gender inflections are attested in several indexation domains (e.g., articles, other adnominal modifiers, predicative expressions, pronouns): (a) information about the gender of a noun is likely to be repeated redundantly in discourse (syntagmatic violation of the Principle of One-Meaning–One-Form); (b) the same word class can take several inflections *depending* on the gender of the noun that is indexed in a given discourse domain (paradigmatic violation of the

---

[14] As mentioned in section 3.2, gender systems with only semantic assignment rules are quite common crosslinguistically, whereas gender systems with only formal assignment rules are almost never encountered.

Principle of One-Meaning–One-Form and Principle of Independence).

Features M1, M2 and CUM are based on an aspect of the typology of gender that falls outside the scope of Audring's work: how grammatical gender interacts with other nominal domains. Two domains are specifically targeted by my metric: number and evaluative morphology (i.e., the morphological encoding of diminutives and augmentatives).[15] M1 and M2 are concerned with interactions at the level of gender assignment whereas CUM has to do with interactions pertaining to the morphosyntactic encoding of gender distinctions on the indexing targets. I suggest that M1 and M2 can be interpreted as a violation of the Principle of Independence, and CUM as a violation of the Principle of One-Meaning–One-Form. Let us discuss these two types of interaction more in detail.

Di Garbo (2014) shows that an important criterion for the classification of gender systems in the African macro-area is to distinguish between *rigid* and *manipulable* gender assignment (for as similar suggestion, see also the study by Heine 1982). In languages with manipulable gender assignment, the gender of a noun can be changed depending on the construal of the noun phrase referent, that is based on pragmatic/discourse constraints. In these languages, there usually are default assignment rules, i.e., rules by which nouns have lexically specified gender values, and *add-on* assignment rules that allow speakers to modify the default meaning of the noun by changing its gender, thus changing the construal of the noun phrase referent. In Di Garbo's sample, manipulable gender assignment is attested in connection with two main uses: (1) to encode variation in the countability properties of nouns (e.g., from uncountable to countable and vice versa), (2) to encode variation in size (diminutive vs. augmentative). In my metric, I refer to the first use of manipulable gender assignment as M1[16] and to the second as M2. M1 is illustrated in example (1) and M2 in example (2). The examples are taken from two Berber languages, Nefusi and Tachawit.[17]

---

[15] The choice of number and evaluative morphology as domains of analysis does not exhaust the whole range of nominal and non-nominal grammatical domains that gender can interact with (among which, for instance, *case* and *definiteness*). However, as shown in section 5, even though far from exhaustive, the metric proposed in this study is able to reveal a good deal of crosslinguistic variation with respect to the complexity of gender and can thus be considered a starting point towards more comprehensive models of interactions of gender with other domains and their effect on gender complexity. For an overview of interactions between domains also involving gender and other systems of nominal classification, see Aikhenvald & Dixon (1998).

[16] Polarity phenomena, whereby polar opposites within a gender and number inflectional paradigm (e.g., masculine singular and feminine plural) have the same type of encoding, do not count as an instance of M1, as defined in this paper. In languages that exhibit polarity (e.g., the Cushitic language Somali), the gender shifts that occur between singular and plural depend on paradigm-specific patterns of exponence and syncretism (often restricted to only a subset of indexing targets), which do not affect the semantic and pragmatic construal of the noun phrase referent with respect to its countability and quantifiability properties (as it happens instead in those languages that I classify as instances of M1). For a discussion of polarity patterns in Somali, see Corbett 1991: 195-197.

[17] The following abbreviations are used in the glossed examples: F = feminine; M = masculine, SG = singular, PL = Plural. The glossing of the examples conforms to the Leipzig Glossing Rules: http://www.eva.mpg.de/lingua/resources/glossing-rules.php

(1) Nefusi (Berber) (Adapted from Beguinot 1942: 32)
   (a)   *ettefâḥ*
          'apples' (masculine, uncountable)

   (b)   *t-attefâḥ-t*
          F-apples-F[SG]
          'one apple'

   (c)   *t-attefâḥ-în*
          F-apples-F.PL
          'apples' (plural)

(2) Tachawit (Berber) (Adapted from Penchoen 1973: 12)
   (a)   *aq-nmuš*
          [M]SG-pot 'pot'

   (b)   *t-aq.nmuš-t*
          F-SG-pot-F
          'small pot'

   (c)   *t-aɣ-nžak-t*
          F-SG-spoon-F
          'spoon'

   (d)   *aɣ-nž*
          [M]SG-spoon
          'big spoon, ladle'

    In example (1) (taken from Nefusi), when the inherently masculine uncountable noun *ettefâḥ* 'apples' is shifted to the feminine gender (as in (1b)), it becomes countable and can be thus regularly pluralized (as in (1c)) (in Berber, feminine gender marking on nouns is circumfixal both in the singular and in the plural). This is an instance of M1. In Tachawit (example (2)), inherently masculine nouns can be shifted to the feminine gender when a diminutive interpretation is intended for the noun phrase referent (as in (2a) and (2b)). Similarly, an inherently feminine noun can be assigned to the masculine gender when an augmentative interpretation is intended for the noun phrase referent (as in (2c) and (2d)). This is an instance of M2.[18] In general M1 and M2 are well attested in the languages of Africa, both in languages with large, non-sex-based gender systems and in languages with smaller sex-based systems. Within my sample, M2 is however more frequent and widely distributed than M1 (for an overview, see Di Garbo 2014: chapters 5 and 6). The possibility of manipulating gender assignment can be seen as piling on top of the default gender assignment rules that

---

[18] A distinction can be made between languages with dedicated diminutive and augmentative genders (as in the Bantu languages), and languages in which there are no diminutive and augmentative genders, but gender shifts between, say, masculine and feminine, are used to encode diminutive and augmentative meanings (as in the Berber languages). This distinction, and its relevance for gender complexity, are not directly addressed by my metric. However, it can at least be observed that, within the sample, languages with dedicated diminutive and augmentative genders are languages with a high number of gender distinctions, which score high with respect to feature GV (see table 2). The relationship between presence of dedicated diminutive and augmentative genders and gender complexity would deserve to be further investigated.

are used in a language. In languages with manipulable gender assignment, gender markers have default and add-on meanings. These add-on meanings are dependent on semantic and pragmatic associations between gender and other grammatical domains, notably countability and size/value. Thus, based on the *Principle of Independence* introduced above, their presence represents an increase in the absolute complexity of gender. Gender assignment is not only given in the lexicon for each and every noun, but it is also subject to change depending on semantic and pragmatic associations with other functional domains.

Feature CUM (cumulative encoding of gender and number on the indexing targets) evaluates the impact that type of exponence of gender and number has on the complexity of gender. I interpret cumulative encoding of gender and number as a violation of the Principle of One-Meaning–One-Form (one morpheme expresses several grammatical meanings). One aspect of the morphosyntactic encoding of gender and number which, at least in the languages of my sample, appears to be strictly related to CUM is the tendency for gender distinctions to be reduced (syncretism) or lost (neutralization) in the context of non-singular number values. In my sample, syncretism and/or neutralization of gender in the context of nonsingular number occurs in 66 out of 84 languages; in nearly all these cases the languages in which syncretism is attested are also languages in which gender and number are encoded cumulatively (see also results in Di Garbo 2014: chapter 5).[19] In principle, gender syncretism and neutralization could be viewed as violations of the Principle of Independence inasmuch as, when they occur, the expression of gender within an inflectional paradigm depends on the number value of a noun. In addition, syncretism and neutralization could be also seen as violations of the Principle of One-Meaning–One-Form, given that two (or more) gender values are conflated into one in the context of non-singular number values. However, as Audring (2016) points out, "[s]yncretism is a multifaceted phenomenon, and whether or not it should be considered a case of simplification or complexification depends on the perspective". In this paper, I treat syncretism in a somewhat agnostic way and exclude it from my complexity metric. More research, I believe, is needed on the relationship between syncretism/neutralization, exponence, and paradigm size before we can assess the effects of syncretism/neutralization on the complexity of gender and related features (e.g., number and case) more confidently.

4.3 Method for computing Gender Complexity Scores

Having defined the features for measuring the absolute complexity of grammatical gender (see table 2), the next step is to establish the values associated with each feature and to convert them into numbers. Towards this aim, I follow Parkvall (2008) who designed a method for computing the grammatical complexity of creoles and non-creole languages on the basis of a set of features taken from the WALS database (Dryer & Haspelmath 2013). Within Parkvall's method, the values of each feature are assigned a number between 0 and 1. Features with three values are converted into the numerical format 0, ½, 1. Similarly, features with five values are converted by Parkvall into the format 0, ¼, ½, ¾, 1. For all the features taken into account in Parkvall's paper, 0 stands for *minimally complex* and 1 for *maximally complex*. The total complexity score for each language is divided by the number of features included for that language. This is done in order to allow languages for which less information is available on a given feature to get average scores comparable to those of the best documented languages. The same procedure is followed in this paper (naturally, features with four values are converted into the numerical format 0, ⅓, ⅔, 1). The feature values and their numerical interpretation are illustrated in table 3.

---

[19] On the relationship between syncretism and cumulative exponence in the domain of case and number inflection, see the studies by Carstairs (1987) and Carstairs & Stemberger (1988).

Feature	Feature Value	Score
GV[20]	Two genders	0
	Three	1/3
	Four	2/3
	Five or more	1
AR	Purely semantic or purely formal assignment	0
	Semantic or formal assignment	1
IND[21]	One	0
	Two	1/3
	Three	2/3
	Four or more	1
CUM	Noncumulative	0
	Partially cumulative	½
	Cumulative	1
M1	Absent	0
	Present	1
M2	Absent	0
	Present	1

Table 3: Gender complexity metric

The composition of the metric is such that the least complex possible gender system is the one that scores zero with respect to all the features of the metric and exhibit the following properties: two gender values, semantic gender assignment, one indexing target, no cumulation with number, no manipulation of gender assignment triggered by number/countability and no manipulation of gender assignment triggered by size. On the other hand, the most complex possible gender system is the one that scores 1 with respect to all the parameters considered in the metric and exhibits the following properties: five or more genders, semantic and formal assignment, four or more indexing targets, cumulation with number, and manipulation of gender assignment triggered by both number/countability and size. In addition, the composition of the metric is such that, with the exception of languages with the highest score (= 1), languages may display the same index value but arrive to it on different paths. In other words, identical gender complexity scores (henceforth GCSs) do not stand for *same type of gender system*.

Before presenting the results of my calculations, it is worth mentioning that, in case of missing features, the index values resulting from the calculations should be taken with caution. In fact, even though average scores (rather than total scores) are used as index values, the index values of languages with missing features cannot be regarded as entirely comparable to the

---

[20] The cut-off point for feature GV is in accordance with the coding conventions for number of gender distinctions proposed by Corbett (2013a).

[21] The coding for feature IND is based on the number of morphosyntactic domains that exhibit gender inflection in a given language. The cut-off point was set at four based on a convenience choice of four main domains of gender inflection: (1) articles (definite/indefinite articles), (2) other adnominal modifiers (including adjectives, demonstrative and possessive modifiers, numerals, quantifiers), (3) predicative expressions, (4) pronouns (including personal pronouns, demonstrative pronouns, possessive pronouns, relative pronouns). This choice was made based on documented crosslinguistic tendencies in the distribution of types of gender indexing targets, as well as in partial overlap with the Agreement Hierarchy proposed by Corbett (1979, 1991, 2006). To these four domains of gender inflection, a category "other" was added in order to account for less prototypical indexing targets (e.g., conjunctions). Within each of the four indexation domains, several word classes may exhibit gender inflection (e.g., personal pronouns and demonstrative pronouns within the pronominal domain); this is not directly addressed by the metric (even though it has been accounted for during data collection). The coding proposed in this study is, of course, only one of the possible ways to model complexity in the domain of gender indexation. For a discussion, see section 7.1.

index values of languages for which all features are equally represented. The mutual comparability between the different domains of gender complexity covered by my metric is discussed in section 6.3.

5. Results

Table 4 illustrates the GCSs of the languages of the sample, which have been calculated based on the method presented in section 4.

Rank	Language	Isocode	GCS	Rank	Language	Isocode	GCS
1	Bandial	bqj	1	8	Gola	gol	0.67
1	Bemba	bem	1	8	Hausa	hau	0.67
1	Bidyogo	bjg	1	9	Awngi	awn	0.61
1	Chiga	cgg	1	9	Hadza	hts	0.61
1	Kagulu	kki	1	9	Moroccan Arabic	ary	0.61
1	Kikuyu	kik	1	9	Nama	naq	0.61
1	Lega	lea	1	9	Naro	nhr	0.61
1	Maasina Fulfulde	ffm	1	9	Sandawe	sad	0.61
1	Mongo-Nkundu	lol	1	9	Standard Arabic	arb	0.61
1	Makaa	mcp	1	9	Tigre	tig	0.61
1	Ndengereko	ndg	1	10	Miya	mkf	0.6
1	Shona	sna	1	11	Male	mdy	0.56
1	Serer	srr	1	11	Wolaytta	wal	0.56
1	Swahili	swh	1	12	Borana-Arsi-Guji Oromo	gax	0.53
1	Timne	tem	1	12	Lishán Didán	trg	0.53
1	Tonga	toi	1	12	Qimant	ahg	0.53
1	Venda	ven	1	12	Rendille	rel	0.53
1	Xoon	nmn	1	12	ǁAni	hnh	0.53
2	Nyanja	nya	0.95	13	Beja	bej	0.5
2	Tunen	baz	0.95	13	Masai	mas	0.5
3	Bafia	ksf	0.83	13	Somali	som	0.5
3	Dibole	bvx	0.83	14	Daasanach	dsh	0.47
3	Eton	eto	0.83	14	Dirasha	gdl	0.47
3	Northern Sotho	nso	0.83	14	Kxoe	xuu	0.47
3	Swati	ssw	0.83	14	Lele	lln	0.47
3	Turkana	tuv	0.83	15	Dizin	mdx	0.45
3	Wamey	cou	0.83	15	Hebrew	heb	0.45
3	Zulu	zul	0.83	15	Gidar	gid	0.45
4	Maltese	mlt	0.78	15	Tsamai	tsb	0.45
4	Noon	snf	0.78	16	Iraqw	irk	0.43
4	Nuclear Wolof	wol	0.78	17	Baiso	bsw	0.42
4	Sɛlɛɛ	snw	0.78	18	Dime	dim	0.39
4	Tswana	tsn	0.78	19	Juǀʼhoan	ktz	0.36
5	Bench	bcq	0.75	19	Kambaata	ktb	0.36
5	Kissi	kss	0.75	20	Dahalo	dal	0.28
6	Karamojong	kdj	0.72	21	Koorete	kqy	0.25
7	Kabyle	kab	0.69	21	Kwadi	kwz	0.25
7	Nafusi	jbn	0.69	22	Lingala, Kinshasa	lin	0.22
7	Tachawit	shy	0.69	23	Bila	bip	0.16
7	Tamasheq, Kidal	taq	0.69	24	Pero	pip	0.12
7	Tamazight, Central	tzm	0.69	25	Mwaghavul	sur	0.08
7	Zenaga	zen	0.69				
8	Amharic	amh	0.67				

Table 4: GCSs of the languages of the sample

The table is divided in two macrocolumns and the GCSs of the individual languages are arranged from highest to lowest. The leftmost columns of each macro-column provide the

rank: languages with the same average complexity score share the same rank. Next to the rank come the language names and their ISO code; the GCS assigned to each language is given in the rightmost columns of the two macro-columns. In appendix C, the GCSs are visualized on the basis of genealogical units. The complexity scores for each of the feature values in the metric, as well the GCSs, are given in appendix B.

Table 4 shows that the highest GCS is 1 and the lowest 0.08. None of the languages of my sample thus gets the lowest possible score, 0 (see section 4.3). The results given in table 4 are also displayed in the graph in figure 1. The X-axis of the histogram displays the range of attested GCSs, whereas the Y-axis shows the distribution of the number of languages per GCS score. The box plot below the histogram provides the distribution of the GCSs per quartiles, with the boldface line in the middle representing the median. The figure shows that half of the languages of my sample have a GCS that ranges roughly from 0.5 to 0.8. In my data sample, high GCSs are substantially more frequent than low GCSs.
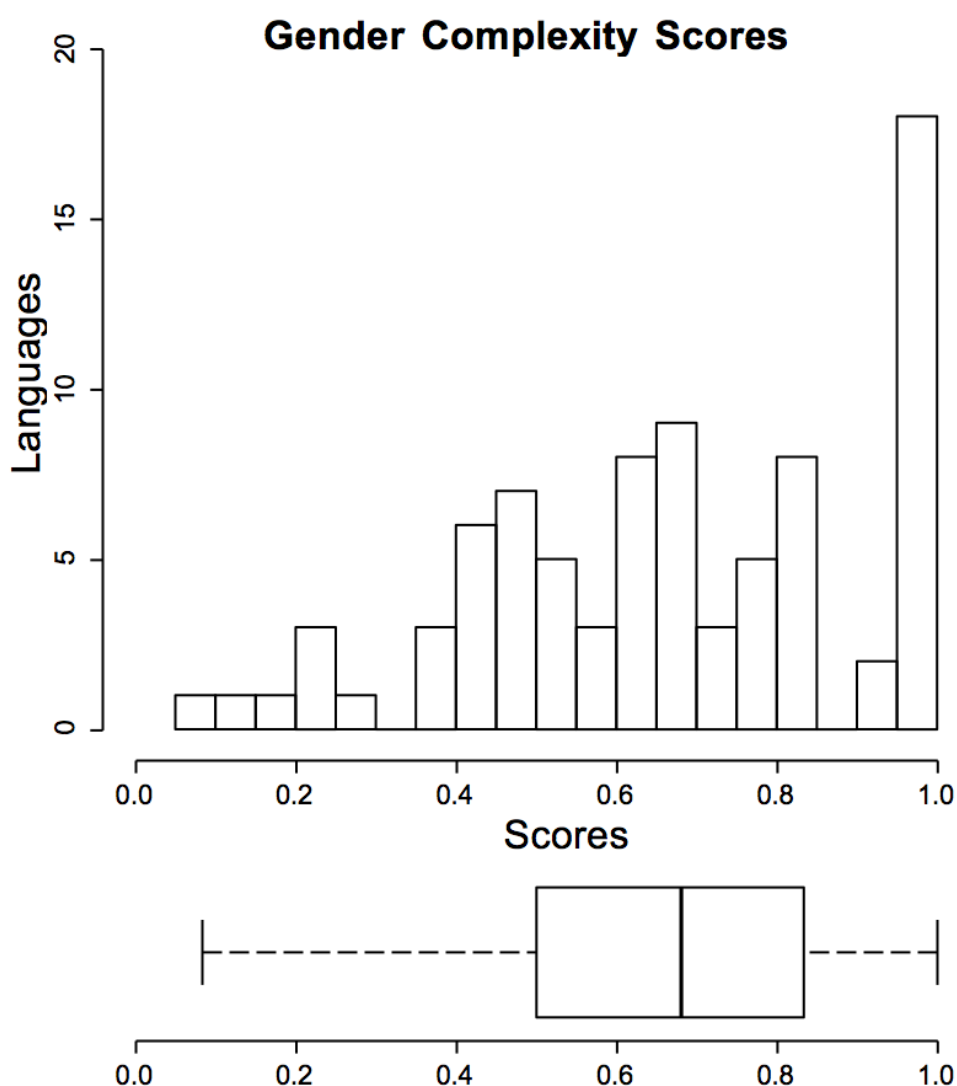
Figure 1: Distribution of the GCSs

The geographical distribution of the GCSs is represented in the map provided in figure 2.
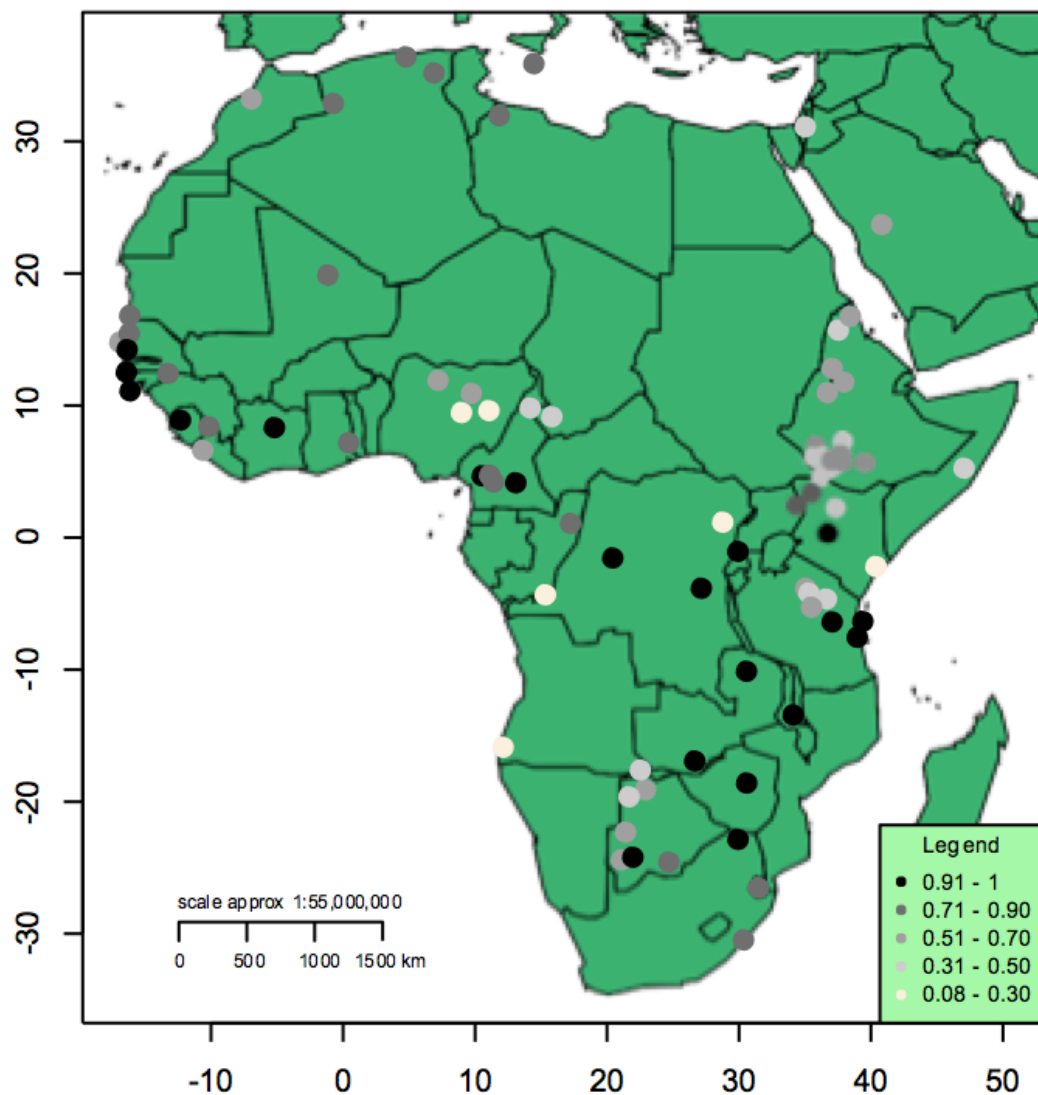
Figure 2: Geographical distribution of the GCSs

The results presented in table 4, figure 1 and 2, as well as in appendix C, are discussed in section 6 based on three main foci:

1. Genealogical distribution of the GCSs
   Languages from the same genealogical units, or spoken within the same areas, tend to have similar or even identical GCSs. In many cases, areal pressure seems to be a relevant factor in explaining the distribution of the outliers.
2. Interdependencies between sets of features: AR, GV, IND
   Purely semantic gender assignment is only found in languages with few genders and poor gender indexation (no directional dependencies between the three features are assumed here).
3. Possible predictors of gender complexity
   Some features in the metric correlate more with each other and seem to have a stronger impact on the GCS than others.

Before moving on to the discussion, I illustrate the procedure followed to calculate the GCSs of two of the sampled languages. For the sake of clarity, I discuss one language for which all features are documented, Turkana (Eastern Nilotic, rank 3 in table 4), and one for which two features are missing, Timne (Mel, rank 1 in table 4).

My classification of the gender system of Turkana is based on Dimmendaal (1983). Turkana has three gender values: Masculine, Feminine and Neuter. It thus gets $\frac{1}{3}$ with respect to the feature GV. Gender assignment is both semantic and formal, and, as such, the value of AR is 1. According to Dimmendaal, gender indexation appears in three domains: articles (definite articles), adnominal modifiers, and pronouns (not the Personal Pronouns). Thus the language gets $\frac{2}{3}$ with respect to the feature IND. In Turkana, gender distinctions are encoded cumulatively with number (CUM = 1). Finally, in Turkana gender shifts can be used to encode variation both in the countability properties of nouns (M1 = 1) and in the size of the noun phrase referent (M2 = 1). In Turkana, when an uncountable masculine or feminine noun is shifted to the Neuter Gender,[22] the resulting meaning is singulative. On the other hand, when countable masculine or feminine nouns are shifted to the Neuter Gender, the resulting meaning is diminutive. To summarize, for Turkana, the values assigned to each feature of the metric are:

$$GV = \tfrac{1}{3}; \; AR = 1; \; IND = \tfrac{2}{3}; \; CUM = 1; \; M1 = 1; \; M2 = 1$$

Applying the formula illustrated in section 4 $[(\tfrac{1}{3} + 1 + \tfrac{1}{3} + \tfrac{2}{3} + 1 + 1 + 1) \div 6]$ the GCS of 0.83 is obtained.

I classify the gender system of Timne based on the description provided by Wilson (1961). Timne has more than five genders and thus gets 1 with respect to the feature GV. Gender assignment is both semantic and formal. Therefore, Timne gets 1 with respect to the feature AR. According to Wilson's description, Timne shows gender indexation on adnominal modifiers, pronouns, predicative expressions. In addition, in Timne, the Indefinite Stabilizer, which is used with indefinite nouns in order to encode non-verbal predication (Wilson 1961: 11), also inflects for gender (this is labeled as "other" in my coding). The language thus gets 1 with respect to IND. Gender and number are encoded cumulatively on the indexing targets (CUM=1). The source does not provide any kind of information about gender shifts, which are, however, rather common phenomena in languages with similar gender systems. The features M1 and M2 cannot be documented for Timne. To summarize, for Timne, the values assigned to each of the metric features are:

$$GV = 1; \; AR = 1; \; IND = 1; \; CUM = 1; \; M1 = -; \; M2 = -$$

Since two features are missing, the sum of the feature values is in this case divided by 4 $[(1+1+1+1) \div 4]$. The GCS of Timne is thus 1.

6. Discussion

6.1 Genealogical and areal biases in the distribution of GCSs

In appendix C, the GCSs presented in table 4 are visualized on the basis of genealogical units. The tables in appendix C show that, in general, closely related languages tend to have the same or very similar GCSs. For instance, all the Berber languages in the sample have a gender complexity score of 0.69. This tendency towards intragenealogical homogeneity in the complexity of gender systems further supports the idea that grammatical gender is a chiefly stable feature in the history of language families (see section 3.1).

---

[22] In this paper, I use capital letters to refer to language-specific categories (e.g., the Neuter Gender in Turkana) and lowercase letters to refer either to a specific marker within a language or to grammatical domains as objects of crosslinguistic comparison (e.g., adjectives and pronouns in Turkana).

Nevertheless, outliers (i.e., languages that exhibit a GCS that is exceedingly higher or lower than what found among closely related languages) are attested in the following genealogical units: Bantu, Chadic, Cushitic, Khoe-Kwadi, Eastern Nilotic, Semitic. I suggest that, at least in some of such cases, the distribution of the outliers can be accounted for by taking into consideration aspects of the social history of the speech communities in question (e.g., geography, number of speakers, number of contact languages, type of language contact, bilingualism, multilingualism). This is however only a preliminary suggestion, which would need to be investigated further in what goes beyond the scope of the present study.

Out of 84 languages, 18 scored 1, with all these being either Bantu, North- Central Atlantic or Mel. Typically, the gender systems of the Bantu and Atlantic type (i.e., North-Central Atlantic and Mel) exhibit features of high complexity: high number of gender distinctions, pervasive gender indexation, manipulability of gender assignment, which is used to express variation in the countability properties of nouns and/or in the size of the noun phrase referents. Those Atlantic and Bantu languages which rank lower than 1 in table 4 have gender systems in which one or more of the above-mentioned features has/have been either weakened or lost. For instance, in 8 of the 23 Bantu languages in the sample – Bafia, Eton, Northern Sotho, Shona, Swati, Tswana, Venda, Zulu – diminutive and augmentative suffixes have grammaticalized from nouns. Of these eight languages, only Venda and, to a lesser extent, Shona combine the use of the diminutive and augmentative suffixes with the uses of the dedicated diminutive and augmentative genders that are characteristic of many Atlantic-Congo languages.[23] In the remaining six languages, the evaluative genders have been lost. As a result, the complexity of the gender systems of these languages is lower than what found in other closely related languages.

Two outliers with respect to the Bantu and Atlantic type of gender system are the Bantu languages Kinshasa Lingala (GCS = 0.22) and Bila (GCS = 0.16). My coding for Kinshasa Lingala, the variety of Lingala spoken in the area of the capital city of the Democratic Republic of Congo, is based on Bokamba (1977) and Meeuwis (2013). Kinshasa Lingala preserves the system of noun class marking which is typical of Bantu languages only on nouns. Meeuwis (2013) rightly refers to this set of singular/plural pairs of nominal prefixes as *inflectional classes*: diachronically, they are a relic of the former Bantu-like gender system, but, synchronically, they merely function as markers of nominal number. The Third Person Pronouns and the Subject Prefixes index the animacy of the noun phrase referent. Based on this account, I classify Kinshasa Lingala as a language with two genders (Animate and Inanimate), semantic gender assignment and two domains of gender indexation (pronominal and predicative). Compared to Makanza Lingala, the northwestern variety of Lingala whose origins go back to the language standardization policies operated by the Scheutist missionaries between 1901 and 1902, and which exhibits a more conservative gender system, the gender system of Kinshasa Lingala is massively reduced. According to Meeuwis (2013: 26), Kinshasa Lingala is the oldest variety and the direct descendant of the Bangala pidgin, which was originally spoken in the Bangala state post (on the northwestern banks of the Congo River) and later on spread northeastward.[24] This variety resisted to the

---

[23] Diminutive and augmentative genders (counted as instances of M2 in my metric) are a distinctive property of the gender systems of many Atlantic- Congo languages. For a description and typological classification of the morpho-syntactic and semantic properties of the evaluative genders in Atlantic-Congo languages and beyond, see Di Garbo 2014, chapter 6.

[24] Bangala is, in turn, the descendant of Bobangi. Bobangi was originally a trade river language spoken on the western part of the Congo River. When the Europeans started using it as a medium of communication, Bobangi underwent a process of massive pidginization with substantial influences from European languages and (mainly) western African languages. The Europeans spread the use of pidginized Bobangi outside its original territory and imposed it as a language of communication in the Bangala Station, where Bangala developed. For more details on

grammatical reforms introduced by the Scheutists, and soon gained both native and second language speakers. The pidginization process from which Lingala originated, as well as the highly multilingual ecology in which the Kinshasa variety developed and expanded, can reasonably explain the patterns of simplification and reduction in the domain of grammatical gender that differentiate this variety from other Bantu languages, on the one hand, and from the standardized variety introduced by the missionaries in the northwestern areas of the Democratic Republic of Congo (Makanza Lingala), on the other (on this account, see also Bokamba 1977, 2009). Similarly to Kinshasa Lingala, Bila has only two genders (the Animate and the Inanimate Gender), semantic assignment rules and poor gender indexation. Differently from Kinshasa Lingala, however, gender indexation in Bila is exclusively internal to the noun phrase and limited to the domain of adnominal modifiers (Kutsch Lojenga 2003: 462). Bila is spoken in the northeastern part of the Democratic Republic of Congo, which is also the northernmost corner of the Bantu-speaking area. The northern part of the Bantu-speaking area is often described as a true borderland between linguistically very diverse communities that have extensive contact with each other. In this area, Bantu speakers are surrounded by speakers of Nilo-Saharan and Ubangi languages (Kutsch Lojenga 2003: 451-452). Due to intense mutual contact, both the Bantu and non-Bantu languages spoken in this area are characterized by massive lexical borrowing as well as by grammatical innovations that are not shared with the respective cognate languages outside the area. The reduced gender system of Bila and other neighboring Bantu languages is one of such area-specific features.

The Semitic languages provide another interesting illustration of a set of genealogically related languages with non-homogeneous GCSs. The highest ranking GCSs within the Semitic sample go to Maltese (0.78) and Amharic (0.67). Moroccan Arabic, Standard Arabic and Tigre have the same complexity score, 0.61. The lowest ranking gender system is found in Hebrew (0.45), whereas Lishán Didán scored 0.53. Interestingly, the highest GCS, 0.78, is scored by Maltese, the Semitic language that stands out for its peculiar history of long-term contact and bilingualism with English, on the one hand, and Romance languages (Italian and Sicilian), on the other. A similarly high GCS goes to Moroccan Arabic, a dialect of Arabic whose history is also characterized by long-term intense contact with Berber languages, French and Spanish (for a case study of complexity of verbal inflection in Moroccan Arabic and other varieties of Arabic, see Kusters 2003). Finally, the history of Modern (Israeli) Hebrew is also intertwined with intricate sociolinguistic dynamics involving processes of creolization, language shift and massive borrowing (see, among others, Doron 2015; Zuckermann 2009).

Two additional examples of outliers are Dahalo, with respect to the other Cushitic languages, and Kwadi, with respect to the Khoe-Kwadi group. Dahalo has a GCS of 0.28, and its gender system has been described by Tosco (1991: 20) as dying out as a result of contact with neighboring Bantu languages. Too little is known about Kwadi, a now extinct language of Angola. Güldemann (2004) describes its gender system as sex-based and pronominal, but not much information is given about mechanisms of gender assignment nor about the use of gender shifts to encode diminutive and augmentative meanings (which is well documented in all the other Khoe-Kwadi languages of the sample).

Finally, the two lowest ranking languages in the complexity rank given in table 4 are the Chadic languages Mwaghavul (GCS = 0.08) and Pero (GCS = 0.12), both of which are spoken in Nigeria. The two languages also qualify as outliers with respect to the other Chadic languages in the sample. Mwaghavul scores 0 with respect to all the features of the complexity metric except for CUM, for which the score is 0.5. There are two genders in Mwaghavul (Masculine and Feminine), gender assignment is semantic and gender indexation

---

the origins and spread of Lingala see Bokamba (2009); Meeuwis (2013).

is only pronominal. Finally, there seems to be no possibility of manipulating gender assignment in the language. With respect to the cumulation parameter, Mwaghavul shows at least some patterns of interaction with number on the indexing targets. The Third Person Human Anaphoric Subject and Object Pronouns encode gender and number cumulatively. On the other hand, the Third Person Non-human Pronoun, *nɔ̄* , encodes neither gender nor number distinctions (Frajzyngier & Johnston 2005). A similar type of system is found in Pero even though, from the description provided by Frajzyngier (1989), it is not entirely clear what type of assignment rules the language has and whether gender assignment is rigid or manipulable. The remaining four Chadic languages in the sample have higher GCSs (between 0.62 – Lele – and 0.45– Gidar). The language-internal and/or socio-historical factors that might account for this distribution should be further investigated.

To summarize, in the languages of my sample, complexity in the domain of grammatical gender tends to be replicated across genealogically related languages. On the other hand, multilingualism, (long term and short term) language contact and second language learning may be seen as possible *disturbance factors* that introduce variation (both in the form of simplification and complexification) in the gender system of a language as opposed to its closest relatives (see also discussion in Trudgill 1999; McWhorter 2001). A systematic account of the effects of sociolinguistic and ecological variables on the complexity of gender falls outside the scope of this paper. In section 7, I put forward a few suggestions on how various aspects of language ecology could be implemented in the study of the grammatical complexity and stability of gender.

6.2 Interdependencies between sets of features: GV and AR, AR and IND

On the basis of the results presented in table 4 an interesting relationship can be observed between the features GV and AR, and AR and IND.

Strictly semantic systems of gender assignment are only found in 8 of the 84 gendered languages within the sample: Bila (Bantu), Dahalo (Cushitic), Dime (South Omotic), Dizin (Dizoid), Kinshasa Lingala (Bantu), Koorete (Ta-Ne-Omotic), Masai (Eastern Nilotic), Mwaghavul (Chadic). All these languages have two gender distinctions, and all but Bila and Kinshasa Lingala have sex-based gender. Within my language sample then, strict semantic gender assignment is only found in languages with two or a maximum of three gender values. Moreover, there seems to be a preference for strictly semantic gender assignment in African languages to be based on cognitively basic oppositions such as *human* vs. *non-human*, *male* vs. *female*, *animate* vs. *inanimate*. It would be interesting to investigate what type of preferences exist, if they exist, in areas of the world where strictly semantic gender assignment is more common.

Finally, it is worth mentioning that the eight languages of my sample with strictly semantic gender assignment all score less than 1 with respect to IND: thus in none of these languages is gender indexation maximally pervasive. These results are in line with a suggestion that was put forward by Audring (2009) with respect to the relationship between pervasiveness of indexation and type of assignment rules. Audring analyzes the assignment rules of a number of pronominal gender systems from different areas of the world, and considers aspects of the diachrony of gender in English and Dutch. She shows that pronominal gender systems – where manifestations of gender throughout the discourse are rather poor – display a strong preference towards strictly semantic assignment rules. Within my language sample, only Mwaghavul (Chadic) has pronominal gender and semantic assignment. However, the remaining five languages with strict semantic assignment score either 1/3 or 2/3 with respect to IND. In line with the expectation voiced in Audring (2009, 2014), these results suggest that when strict semantic gender assignment is found in non-pronominal gender systems, gender indexation is still not maximally pervasive. In other words, semantic assignment seems to

generally tolerate lower amount of formal marking.

6.3 Some features may be stronger predictors of gender complexity than others

As discussed in section 2.2, a major issue when investigating grammatical complexity is how to quantify the contribution that the individual features of a metric bring to the overall complexity score (what Miestamo 2006b, 2008 refers to as *the problem of comparability*). Given that it is extremely difficult to measure the relative weight of the individual features of a complexity metric, as well as to establish the number and type of features to be included in a metric, complexity metrics cannot be interpreted as uncontroversial and exhaustive measurements, but rather as tools to detect and describe tendencies in the complexity of a grammatical domain with respect to a selection of relevant features (for a similar discussion in a study of complexity in nominal plural allomorphy, see also Dammel & Kürschner 2008). I would like to suggest here that one way of indirectly investigating the behavior of a complexity metric is to correlate the individual features with each other. In order to do so with my own metric, I calculated the Squared Spaerman rank correlation coefficients between the individual features of the metric. The results are represented in the graph in figure 3.

Figure 3 is organized as follows. The individual features of the metric are displayed both horizontally and vertically. In this way, correlations coefficients between pairs of features can be read both row-wise and column-wise. Correlation coefficients are visualized according to a color scale whereby white stands for *no correlation* and gray for *high correlation*. The gray diagonal area that cuts across the two halves of the figure represents correlation coefficients between pairs of the same features (that is, CUM with CUM, M2 with M2, etc.). These gray boxes correspond to a correlation coefficient that equals to 1 since each feature obviously has the highest correlation with its own copy. These results are thus not relevant to the analysis. With respect to correlations between pairs of different features, the figure shows that the highest correlation coefficients are found between IND and M1 (= 0.353), GV and IND (= 0.295) and GV and M1 (= 0.261).
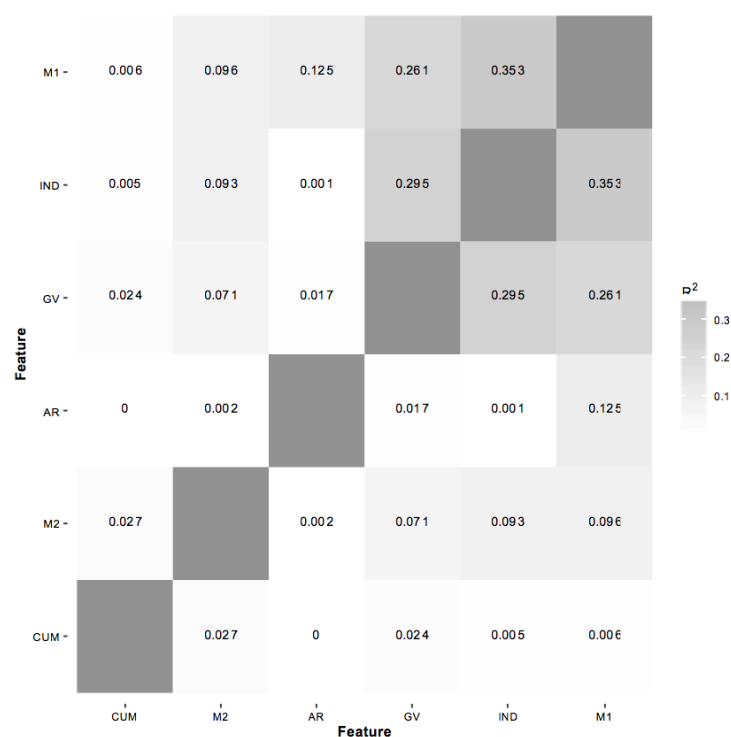

Figure 3: Correlation coefficients between the features of the metric

The correlation coefficients between IND and M1, and, to a lesser extent, between GV and M1 can be interpreted as follows. In the languages of my sample, the possibility of manipulating gender assignment to encode variation in the countability properties of nouns goes hand in hand with the presence of very pervasive gender indexation or, to a lower degree, high number of gender values. M1 is not widely distributed across the language sample. It is only found in Bantu (with the exception of Bila and Kinshasa Lingala), North-Central Atlantic, Berber, a subset of the Semitic languages, and in the Eastern Nilotic language Turkana. In a way then, both the distribution of M1 and its correlation coefficients with IND and GV suggest that M1 is a very special property of gender systems, which can only be found in systems with a high amount of formal marking (IND) and/or a high number of gender distinctions (GV). On the contrary, the results show that M2, that is, manipulation of gender assignment to express diminutive and augmentative meanings, has extremely low correlation coefficients with both IND and GV as well as with all the other features of the metrics.

As mentioned above, GV and IND exhibit a relatively high correlation coefficient, 0.295. This result supports Audring's (2014) argument, whereby a high number of gender distinctions is likely to be found in languages with pervasive indexation (see section 3.2).

Moreover, figure 3 shows that AR has extremely low correlation coefficients with all the features of the metric. These results might depend on the fact that only 8 of the 84 sampled languages have semantic gender assignment. In other words, nearly all the languages of the sample behave similarly with respect to this parameter. It would be interesting to investigate the behavior of this feature in areas of the world where semantic gender assignment is more frequent and compare it with the results from Africa. Finally, equally low correlations are found with the feature CUM.

One question that is worth asking is whether the correlation coefficients presented in figure 3 can tell us anything about which of these features is the best predictor of the GCS of each language. Since the GCS is the averaged sum of the values that a language takes for each feature in the metric, the features that show the highest correlations with each other (M1, IND and GV) can be expected to be those which also have a stronger impact on the final score. This can be verified by examining the associations between the independent variables (the features in the metric) and the dependent variable (the GCS) in a purely descriptive way, that is, by stratifying our dependent variable, the GCSs, according to the potential predictors, the individual features in the metric (Harrell 2001: 125). This is shown in figure 4.
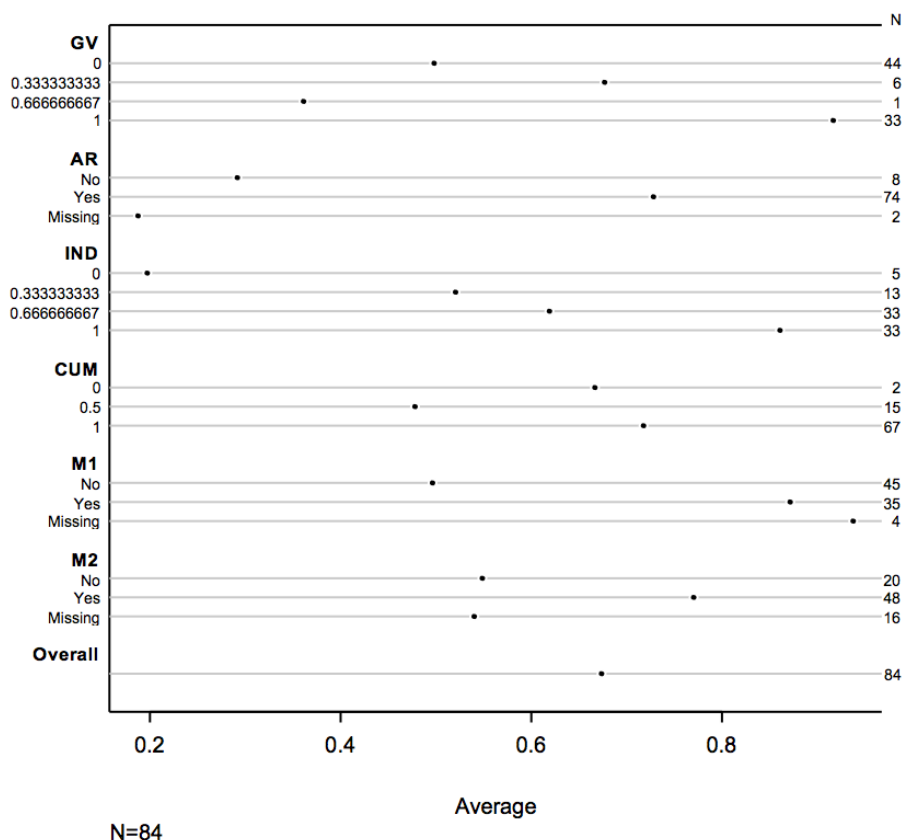
Figure 4: GCSs (Average) stratified according to feature values

Figure 4 is organized as follows. The GCSs are displayed on the X-axis. The left Y-axis represents the values assigned to each feature in the metric; the right Y-axis shows the number of languages in the sample where each of the feature values is found. The black dots represent the mean of the GCSs that languages displaying a certain feature value have. For instance, it shows that languages that score ⅓ (0.3333333333) with respect to GV have a GCS which, on average, ranges between 0.6 and 0.8. The black dots thus allow us to see which of the features and feature values can trigger the highest GCSs in the languages of the sample. As hypothesized based on the correlation coefficients shown in figure 3, in the languages of my sample, the highest scores in GV, IND and M1, trigger higher GCSs. With respect to GV, the figure shows that the impact of the different feature values on the GCSs grows from 0 to 1/3, drastically drops at 2/3 and grows again at 1. This is likely to be an effect of the fact that only one language within my sample has four gender distinctions, Ju|'hoan (Kxa). Ju|'hoan has a GCS of 0.36, which is one of the lowest scores in my language sample.

To summarize, even though the quantitative analysis applied to the data does not provide a solution to the problem of comparability, it provides valid tools for describing the behavior of the complexity metric with respect to the data-set investigated in this paper. Provided that my metric is a good measure for (at least some aspects of) gender complexity, the results suggest that GV, IND and M1 are the features which correlate more strongly with each other and which seem to have the strongest impact on the final complexity scores of the languages of the sample.

7. Summary and concluding remarks

The aim of this paper was to contribute to the debate on the empirical study of grammatical complexity by proposing a set of theoretical principles and methodological

tools that can be used to investigate the complexity of grammatical domains in a typological perspective. The study focused on one grammatical domain, gender, which was chosen in virtue of its well known association with morphosyntactic complexity (inflection and indexation), diachronic stability and areal persistence.

With respect to *theoretical assumptions*, linguistic complexity was here conceived of in terms of number of parts/description length of a given system. It was argued that typological complexity metrics should focus on individual grammatical domains and that the complexity of a given domain should be evaluated against three principles: the Principle of One-Meaning–One-Form, the Principle of Fewer Distinctions, and the Principle of Independence.

With respect to *methodology*, the study followed a sampling procedure that exploits areal and genealogical biases with the purpose of investigating if, and to which extent, typological distributions concerning the complexity of gender systems are genealogically and areally entrenched. Finally, the study provided an empirical illustration of how complexity metrics may be designed and implemented quantitatively. This was done by expanding on the dimensions of gender complexity suggested by Audring (2014) and converting them into a set of features with measurable values. Complexity scores for each of the sample languages were then calculated on the basis of a method introduced by Parkvall (2008).

In section 7.1 and 7.2, I evaluate the main contributions of the investigation with respect to: (a) the complexity metric proposed and (b) the results obtained in the study.

7.1 Evaluation of the complexity metric

The metric designed for this study consisted of six features and assessed the complexity of grammatical gender based on the following parameters: number of gender distinctions, gender assignment, patterns of indexation, interactions with two other nominal domains – number and evaluative morphology – as reflected via gender assignment (manipulation of gender assignment) and type of exponence of gender on the indexation targets (cumulation with number). The six features are not to be understood as an exhaustive inventory of complexity parameters for gender, but as a first attempt to translate a set of crosslinguistically documented properties of gender systems into indexes of complexity. Here I make some suggestions about how the metric could be further improved.

First of all, the metric proposed in this study does not include gender marking on nouns (e.g., presence vs. absence of overt gender, type of exponence of gender on nouns) as one of the dimensions for assessing the complexity of a gender system. This choice was motivated by the idea that, in order to investigate the complexity of gender, one should first look at the domain of encoding that is most definitional of this morphosyntactic feature, i.e., indexation (there is no gender if there is no indexation). Nevertheless, understanding how overt gender marking on nouns affects the overall complexity of a gender system is a promising area to explore in further studies of the complexity of gender. One suggestion that is put forward by Audring (2016) is that, based on the Principle of One-Meaning–One-Form (or Principle of Transparency in her own terminology), covert gender systems are more complex than overt gender systems because in covert gender systems, nouns fail to mark a morphosyntactic feature that they inherently carry.

Second, further research is particularly needed to improve the analysis of gender indexation patterns. In my metric, the amount of gender indexation per each of the sample languages is established by counting the morphosyntactic domains in which gender marking occurs in a language. As explained in section 4.3, footnote 21, this is done by identifying the word classes that carry gender inflection and by ascribing them to one of five possible codings for indexation domains (articles, other adnominal modifiers, predicative expressions, pronouns, and others). Thus feature IND provides a rough count of how pervasive gender indexation is in a language, but does not allow us to immediately verify whether, for

instance, "one indexing domain" means "only pronominal" or "only adnominal modification", or how many word classes inflect for gender within each of the relevant domains (e.g., within the pronominal domain, only personal pronouns or personal pronouns and demonstrative pronouns). Moreover, gender indexes are identified on the basis of a set of distinguishable functions (e.g, modification in the case of adjectives, predication in the case of verbs etc.). Two functionally different indexes (e.g., definite articles and demonstrative pronouns/modifiers) can have the same formal realization in one language. However, the metric does not account for the implications of these patterns of identity of forms on the complexity of individual gender systems. On a more general level, accounting for the difference between gender systems in which the gender indexing targets have the same formal realization and those in which indexing targets are formally distinct might be crucial, for instance, when investigating the relationship between *complexity* and *difficulty* in the domain of grammatical gender. This line of research falls, however, outside the scope of the present investigation. In addition, the metric does not directly account for the frequency of gender marking in discourse, an issue that would be also worth exploring when examining the relationship between complexity and difficulty.

Finally, even though the metric allows for exploring interactions between gender and other nominal features, the inventory of possible interactions is far from exhaustive, mainly because restricted to only two domains (number and evaluative morphology). Further research is needed on each of these issues, whose relevance has also been recently discussed by Audring (2016).

7.2 Evaluation of the results and prospects for future research

The gender systems of the African languages sampled for this study are generally associated with high degrees of complexity (see section 5). In addition, the results show that the complexity of grammatical gender is likely to be replicated across genealogically related languages. If these results are interpreted in terms of stability, one could speculate that, at least in this area of the world, not only are noncomplex gender systems infrequent, but that they also represent diachronically unstable stages in the history of languages. However, as discussed in section 6.1, some outliers were found in almost all the genealogical groupings represented in the sample. In many such cases, the outlier languages tend to stand out from closely related languages because of rather distinctive socio-historical factors: (1) high degree of multilingualism/nonnative acquisition (e.g., Kinshasa Lingala and Modern Hebrew), (2) intense long-term contact and bi- or multilingualism with languages lacking gender or displaying different types of gender systems (e.g., Bila, Dahalo, Maltese). These results suggest that a grammatical feature like gender, which appears to be rather stable when looking at genealogical and areal distributions at the macro-level, in fact exhibits striking patterns of variation when family-internal comparisons are carried out at the micro-level. In this sense, the study shows that investigating how related languages differ in complexity with respect to specific domains of grammar can be a promising way to explore the stability of these domains.

The results of the study also point to the necessity of integrating language ecology[25] in the typological study of the complexity of grammatical domains. Only by implementing socio-historical factors as variables of our complexity metrics can we explore the extent to which these factors contribute to grammatical complexification and/or simplification crosslinguistically. I would like to argue here that integrating language ecology in the crosslinguistic study of linguistic complexity is central to the development of *sociolinguistic*

---

[25] Haugen (1972: 325) defines language ecology as the "the study of interactions between any given language and its environment."

typology (Trudgill 2011) taken both as a method and a theory of research on linguistic diversity (for a similar approach to the study of the social determinants of linguistic complexity see also Lupyan & Dale 2010 and their *Linguistic Niche Hypothesis*, whereby the distribution of linguistic complexity is conceived of as due, at least in part, to the different social environments in which languages are learned and used). By implementing methods that systematically assess the intersections between ecological profiles and the complexity of grammatical domains, an ecology-informed approach to the typological study of linguistic complexity may also contribute to reducing, and ultimately overcoming, the gap between *relative* and *absolute* approaches (see section 2.1).

In conclusion, the metric and the methodology proposed in this study are, in many respects, only a preliminary and far from exhaustive attempt at assessing the complexity of grammatical gender within and across languages. Nevertheless I hope to have shown that this attempt is not just a sterile exercise in determining how "rich" languages can be with respect to a specific domain of grammar, but rather a promising tool for exploring the distribution of linguistic diversity and understanding the internal and external dynamics that constraint the raise and spread of this diversity.

References

Aikhenvald, Alexandra. 2003. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.

Aikhenvald, Alexandra & Robert M. W. Dixon. 1998. Dependencies between grammatical systems. *Language* 74(1). 56–80.

Amha, Azeb. 2012. Omotic. In Zygmunt Frajzyngier & Erin Shay (eds.), *The Afroasiatic languages*, 423–504. Cambridge: Cambridge University Press.

Audring, Jenny. 2009. Gender assignment and gender agreement: Evidence from pronominal gender languages. *Morphology* 18. 93–116.

Audring, Jenny. 2014. Gender as a complex feature. *Language Sciences* 43. 5–17. [Special issue: Exploring grammatical gender].

Audring, Jenny. 2016. Calibrating complexity. *Language Sciences* Special issue.

Bakker, Dik. 2011. Language sampling. In Jae Jung Song (ed.), *Handbook of linguistic typology*, 100–127. Oxford: Oxford University Press.

Beguinot, Francesco. 1942. *Il berbero di Nefûsi di Fassâto*. Roma: Istituto per l'Oriente.

Bickel, Balthasar. 2013. Distributional biases in language families. In Alan Timberlake, Johanna Nichols, David A. Peterson, Balthasar Bickel & Lenor A. Grenoble (eds.), *Language typology and historical contingency: In honor of Johanna Nichols*, 415–443. Amsterdam: John Benjamins.

Bokamba, E. 2009. The spread of Lingala as a lingua franca in the Congo Basin. In Fiona McLaughlin (ed.), *The languages of urban Africa*, 50–70. London: Continuum.

Bokamba, Eyamba. 1977. The impact of multilingualism on language structures: the case of Central Africa. *Anthropological Linguistics* 19. 181–202.

Carstairs, Andrew. 1987. *Allomorphy in inflection*. London: Croom Helm.

Carstairs, Andrew & Joseph Paul Stemberger. 1988. A processing constraint on inflectional homonymy. *Linguistics* 26. 601–617.

Contini-Morava, Ellen & Marcin Kilarski. 2013. Functions of nominal classification. *Language Sciences* 40. 263–299.

Corbett, Greville. 1979. The agreement hierarchy. *Journal of Linguistics* 15. 203–224.

Corbett, Greville. 1991. *Gender*. Cambridge: Cambridge University Press.

Corbett, Greville. 2006. *Agreement*. Cambridge: Cambridge University Press. Corbett,

Greville. 2012. *Features*. Cambridge: Cambridge University Press.

Corbett, Greville. 2013a. Number of genders. In Matthew Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Max Planck Digital Library, chapter 30. Available online at: http://wals.info/chapter/30. Accessed on 2014-02-14.

Corbett, Greville. 2013b. Sex-based and non-sex-based gender systems. In Matthew Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Max Planck Digital Library, chapter 31a. Available online at: http://wals.info/chapter/31. Accessed on 2014-02-14.

Corbett, Greville. 2013c. Systems of gender assignment. 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/32. Accessed on 2014-02-14.

Croft, William. 2001. *Radical construction grammar*. Oxford: Oxford University Press. Croft, William. 2003. *Typology and universals*. Cambridge: Cambridge University Press.

Croft, William. 2013. Agreement as anaphora, anaphora as coreference. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: studies in memory of Anna Siewierska*, 107–129. Berlin: Mouton de Gruyter.

Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.

Dahl, Östen. 2011. Grammaticalization and linguistic complexity. In Heiko Narrog & Bernd Heine (eds.), *The Oxford handbook of grammaticalization*, 153–162. Oxford: Oxford University Press.

Dammel, Antje & Sebastian Kürschner. 2008. Complexity in nominal plural allomorphy. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 243–262. Amsterdam: John Benjamins.

Di Garbo, Francesca. 2014. *Gender and its interaction with number and evaluative morphology: An intra- and intergenealogical typological survey of Africa*. Stockholm: Department of Linguistics, Stockholm University dissertation.

Dimmendaal, Gerrit. 1983. *The Turkana language*. Dordrecht: Foris Publications.

Doron, Edit (ed.). 2015. *Language contact and the development of Modern Hebrew*. Leiden: Brill.

Dryer, Matthew. 1989. Large linguistic areas and language sampling. *Studies in Language* 13. 257–292.

Dryer, Matthew & Martin Haspelmath (eds.). 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info, Accessed on 2014-02-14.

Foley, W. A. & R. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Frajzyngier, Zygmunt. 1989. *A grammar of Pero*. Berlin: Dietrich Reimer Verlag. Frajzyngier, Zygmunt & Eric Johnston. 2005. *A grammar of Mina*. Berlin: Mouton de Gruyter.

Greenberg, Joseph. 1978. How does a language acquire gender markers? In Joseph Greenberg, Charles Ferguson & Edith Moravcsik (eds.), *Universals of human language*, vol. 3: Word structure, 47–92. Stanford: Stanford University Press.

Güldemann, Tom. 2004. Reconstruction through 'de-construction': the marking of person, gender and number in the Khoe family and Kwadi. *Diachronica* 21. 251–306.

Harrell, Frank E. 2001. *Regression modeling strategies*. New York: Springer.

Haugen, Einar. 1972. The ecology of language. In Answar Dil (ed.), *The ecology of language: Essays by Einar Haugen*, 325–339. Stanford: Stanford University Press.

Heath, Jeffrey. 1975. Some functional relationships in grammar. *Language* 51. 89–104.

Heine, Bernd. 1982. African noun class systems. In H. Seiler & C. Lehmann (eds.),

Apprehension: Das sprachliche Erfassen von Gegenständen, 189–216. Tübingen: Gunter Narr Verlag.

Hockett, Charles F. 1958. *A course in modern linguistics*. New York: Macmillan.

Kilarski, Marcin. 2013. *Nominal classification: A history of its study from the classifcal period to the present*. Amsterdam: John Benjamins.

Killian, Don. 2015. *Topics in Uduk phonology and morphosyntax*. Helsinki: Department of World Cultures, African Studies dissertation.

Kusters, Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflections*. Utrecht: LOT: University of Leiden dissertation.

Kutsch Lojenga, Constance. 2003. Bila (D 32). In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 450–474. London: Routledge.

Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PLOS one* 5(1). 1–10.

Maslova, Elena. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4-3(3). 307–333.

McWhorter, John. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5. 125–166.

Meeuwis, Michael. 2013. Lingala. In Susanne Michaelis, Philipe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The survey of pidgin and creole languages*, vol. III, Contact languages based on languages from Africa, Asia, Australia and the Americas, 25–33. Oxford: Oxford University Press.

Miestamo, Matti. 2006a. On the complexity of standard negation. In Mickael Suominen, Antti Arppe, Anu Airola, Orvokki Heinämaki, Matti Miestamo, Urho Määttä, Jussi Niemi, Kari K. Pitkänen & Kaius Sinnemäki (eds.), *A man of measure: festschrift in honour of Fred Karlsson on his 60th birthday [Special supplement to SKY Journal of Linguistics 19]*, 345–356. Turku: The Linguistic Association of Finland.

Miestamo, Matti. 2006b. On the feasibility of complexity metrics. In Krista Kerge & Maria-Maren Sepper (eds.), *Finest Linguistics. Proceedings of the Annual Finnish and Estonian Conference of Linguistics, Tallin, May 6–7, 2004*, 11–26. Tallin: TLÜ.

Miestamo, Matti. 2008. Grammatical complexity in a cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 23–41. Amsterdam: John Benjamins.

Miestamo, Matti, Kaius Sinnemäki & Fred Karlsson (eds.). 2008. *Language complexity: Typology, contact, changes*. Amsterdam: John Benjamins.

Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.

Nichols, Johanna. 2003. Diversity and stability in language. In Brian Joseph & Richard Janda (eds.), *The handbook of historical linguistics*, 283–310. Oxford: Blackwell.

Nichols, Johanna. 2009. Linguistic complexity: a comprehensive definition and survey. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 110–125. Oxford University Press.

Nordhoff, Sebastian, Harald Hammarström, Robert Forkel & Martin Haspelmath. 2013. *Glottolog 2.2*. Max Planck Institute for Evolutionary Anthropology. Available online at http://glottolog.org, Accessed on 2015-09-17.

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*, 265–285. Amsterdam: John Benjamins.

Penchoen, Thomas. 1973. *Etude sintaxique d'un parler berbère (Ait-Frah de l'Aurès)*. Napoli: Centro di studi magrebini.

Sinnemäki, Kaius. 2011. *Language universals and linguistic complexity*. University of

Helsinki: General Linguistics, Department of Modern Languages dissertation.

Tosco, Mauro. 1991. *A grammatical sketch of Dahalo*. Hamburg: Helmut Buske. Trudgill, Peter. 1999. Language contact and the function of linguistic gender. *Poznań Studies in Contemporary Linguistics* 35. 133–152.

Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. New York: Oxford University Press.

Wilson, William A. A. 1961. *An outline of the Temne language*. London: School of Oriental and African Studies.

Zuckermann, Ghil'ad. 2009. Hybridity versus revivability: Multiple causation, forms and patterns. *Journal of Language Contact* 2. 40–67.

Appendix

A. The Language Sample

Languages are listed alphabetically. The language names are followed by the ISO codes, and the names of the genealogical units that each language is assigned to in Glottolog (Nordhoff et al. 2013), as of September, 2015.

Language	ISO	Genealogical Unit
Amharic	amh	Afro-Asiatic, Semitic
Awngi	awn	Afro-Asiatic, Cushitic
Bandial	bqj	Atlantic-Congo, North-Central Atlantic
Bafia	ksf	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Baiso	bsw	Afro-Asiatic, Cushitic
Beja	bej	Afro-Asiatic, Cushitic
Bemba	bem	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Bench	bcq	Ta-Ne-Omotic
Bila	bip	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Borana-Arsi-Guji Oromo	gax	Afro-Asiatic, Cushitic
Bidyogo	bjg	Atlantic-Congo, North-Central Atlantic
Chiga	cgg	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid,Bantu
Daasanach	dsh	Afro-Asiatic, Cushitic
Dahalo	dal	Afro-Asiatic, Cushitic
Dibole	bvx	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Dime	dim	South Omotic
Dirasha	gdl	Afro-Asiatic, Cushitic
Dizin	mdx	Dizoid
Eton	eto	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Gidar	gid	Chadic
Gola	gol	Atlantic-Congo, Mel
Hadza	hts	Isolate
Hausa	hau	Chadic
Hebrew	heb	Afro-Asiatic, Semitic
Iraqw	irk	Afro-Asiatic, Cushitic
Ju|'hoan	ktz	Kxa
Kabyle	kab	Afro-Asiatic, Berber
Kagulu	kki	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Kambaata	ktb	Afro-Asiatic, Cushitic
Karamojong	kdj	Nilotic, Eastern Nilotic
Kikuyu	kik	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Kissi	kss	Atlantic-Congo, Mel
Koorete	kqy	Ta-Ne-Omotic
Kwadi	kwz	Khoe-Kwadi
Kxoe	xuu	Khoe-Kwadi
Lega	lea	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Lingala (Kinshasa)	lin	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Lele	lln	Chadic
Lishán Didán	trg	Afro-Asiatic, Semitic
Masai	mas	Nilotic, Eastern Nilotic
Maasina Fulfulde	ffm	Atlantic-Congo North-Central Atlantic
Makaa	mcp	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Male	mdy	Ta-Ne-Omotic
Maltese	mlt	Afro-Asiatic, Semitic
Miya	mkf	Afro-Asiatic, Chadic
Mongo-Nkundu	lol	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Moroccan Arabic	ary	Afro-Asiatic, Semitic
Mwaghavul	sur	Afro-Asiatic, Chadic
Nafusi	jbn	Afro-Asiatic, Berber

Nama	naq	Khoe-Kwadi
Naro	nhr	Khoe-Kwadi
Ndengereko	ndg	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Noon	snf	Atlantic-Congo, North-Central Atlantic
Northern Sotho	nso	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Nyanja	nya	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Pero	pip	Afro-Asiatic, Chadic
Qimant	ahg	Afro-Asiatic, Cushitic
Rendille	rel	Afro-Asiatic, Cushitic
Sandawe	sad	Isolate
SElEE (spelled *Selee* in Glottolog)	snw	Atlantic-Congo, Volta-Congo, Kwa
Serer	srr	Atlantic-Congo, North-Central Atlantic
Shona	sna	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Somali	som	Afro-Asiatic, Cushitic
Standard Arabic	arb	Afro-Asiatic, Semitic
Swati	ssw	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Swahili	swh	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Tachawit	shy	Afro-Asiatic, Berber
Tamasheq (Kidal)	taq	Afro-Asiatic, Berber
Tamazight (Central Atlas)	tzm	Afro-Asiatic, Berber
Tigre	tig	Afro-Asiatic, Semitic
Timne	tem	Atlantic-Congo, Mel
Tonga	toi	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Tsamai	tsb	Afro-Asiatic, Cushitic
Tswana	tsn	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Tunen	baz	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Turkana	tuv	Nilotic, Eastern Nilotic
Venda	ven	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
Wamey	cou	Atlantic-Congo, North-Central Atlantic
Wolaytta	wal	Ta-Ne-Omotic
Wolof (Nuclear)	wol	Atlantic-Congo, North-Central Atlantic
Zenaga	zen	Afro-Asiatic, Berber
Zulu	zul	Atlantic-Congo, Volta-Congo, Benue-Congo, Bantoid, Bantu
ǁAni	hnh	Khoe-Kwadi
Xoon	nmn	Tuu

B. Complexity scores for the individual features in the metric

Table 6 shows how each of the sampled languages scored with respect to the features of the complexity metric. Unlike in table 4, where the GCSs are rounded up to numbers with two decimal places, unrounded figures are provided in table 6. The data are ordered alphabetically based on the ISO codes of the sampled languages. See table 5 for the correspondent language names.

Table 6: Complexity scores

ISO	GV	AR	IND	CUM	M1	M2	GCS
ahg	0	1	2/3	1	0		0.533333333
amh	0	1	1	1	0	1	0.666666667
arb	0	1	2/3	1	1	0	0.611111111
ary	0	1	2/3	1	1	0	0.611111111
awn	0	1	2/3	1	0	1	0.611111111
baz	1	1	2/3	1	1	1	0.944444445
bcq	1/3	1	2/3	1	0	1	0.75
bej	0	1	1	0	0	1	0.5
bem	1	1		1	1	1	1
bjg	1	1	1	1	1	1	1
bip	0	0	0	1	0	0	0.166666667
bqj	1	1	1	1	1	1	1
bsw	0	1	1	1/2	0	0	0.416666667
bvx	1	1	1	1	1	0	0.833333333
cgg	1	1	1	1	1	1	1
cou	1	1	1	0	1	1	0.833333333
dal	0	0	2/3	1	0	0	0.277777778
dim	0	0	1/3	1	0	1	0.388888889
dsh	0	1	1/3	1/2	0	1	0.472222222
eto	1	1	1	1	1	0	0.833333333
ffm	1	1	1	1	1	1	1
gax	0	1	2/3	1	0		0.533333333
gid	0	1	2/3	1	0	0	0.444444445
gdl	0	1	1/3	1	0		0.466666667
gol	1	1	1/3	1	0		0.666666667
hau	0	1	1	1	0	1	0.666666667
heb	0	1	2/3	1	0	0	0.444444445
hnh	1/3	1	1/3	1	0		0.533333333
hts	0	1	2/3	1	0	1	0.611111111

ISO	GV	AR	IND	CUM	M1	M2	GCS
irk	0	1	2/3	1/2	0		0.433333333
jbn	0	1	2/3	1/2	1	1	0.694444445
kab	0	1	2/3	1/2	1	1	0.694444445
kdj	1/3	1	1	1	0	1	0.722222222
kik	1	1	1	1	1	1	1
kki	1	1	1	1	1	1	1
kqy	0	0	2/3	1/2	0	1	0.25
ksf	1	1	1	1	1	0	0.833333333
kss	1	1	1/3	1		0	0.75
ktb	0	1	2/3	1/2	0	0	0.361111111
ktz	2/3	1	0	1/2	0	0	0.361111111
kwz	0		0	1	0		0.25
lea	1	1	1	1	1	1	1
lin	0	0	1/3	1	0	0	0.222222222
lln	0	1	1/3	1	0		0.466666667
lol	1	1	1	1	1	1	1
mas	0	0	1	1	0	1	0.5
mcp	1	1	1	1	1	1	1
mdx	0	0	2/3	1	0	1	0.444444445
mdy	0	1	1/3	1	0	1	0.555555556
mkf	0	1	1	1	0		0.6
mlt	0	1	2/3	1	1	1	0.777777778
naq	1/3	1	1/3	1	0	1	0.611111111
nhr	1/3	1	1/3	1	0	1	0.611111111
ndg	1	1	1	1	1	1	1
nmn	1	1	1	1			1
nso	1	1	1	1	1	0	0.833333333
nya	1	1	2/3	1	1	1	0.944444445
pip	0		0	1/2	0		0.125
rel	0	1	2/3	1	0		0.53333333
sad	0	1	2/3	1	0	1	0.611111111
shy	0	1	2/3	1/2	1	1	0.694444445
sna	1	1	1	1	1	1	1
snf	1	1	2/3	1	0	1	0.777777778
snw	1	1	2/3	1	0	1	0.777777778
som	0	1	1	1/2	0		0.5

Table 6: (continued)

ISO	GV	AR	IND	CUM	M1	M2	GCS
srr	1	1	1	1		1	1
ssw	1	1	1	1	1	0	0.833333333
sur	0	0	0	1/2	0	0	0.083333333
swa	1	1	1	1	1	1	1
taq	0	1	2/3	1/2	1	1	0.694444445
tem	1	1	1	1			1
tig	0	1	2/3	1	0	1	0.611111111
toi	1	1	1	1	1	1	1
trg	0	1	2/3	1	0		0.533333333
tsb	0	1	2/3	1	0	0	0.444444445
tsn	1	1	2/3	1	1	0	0.777777778
tuv	1/3	1	2/3	1	1	1	0.833333333
tzm	0	1	2/3	1/2	1	1	0.694444445
ven	1	1	1	1	1	1	1
wal	0	1	1/3	1	0	1	0.555555556
wol	1	1	2/3	1	0	1	0.777777778
xuu	0	1	1/3	1	0		0.466666667
zen	0	1	2/3	1/2	1	1	0.694444445
zul	1	1	1	1	1	0	0.833333333

C. GCSs per genealogical units

In the following, GCSs are visualized on the basis of genealogical units. The genealogical units that are represented by one language only are not included in the appendix. These are: Dizoid (represented by Dizi), Hadza (isolate), Kxa (represented by Ju|'hoan), Kwa (represented by SElEE), Sandawe (isolate), South Omotic (represented by Dime), Tuu (represented by !Xoo). The GCSs of these languages are given in table 4.

Table 7: Bantu

ISO	Language	GCS
baz	Tunen	0.944444445
bem	Bemba	1
bip	Bila	0.166666667
bvx	Dibole	0.777777778
cgg	Chiga	1
eto	Eton	0.833333333
kik	Gikuyu	1
kki	Kagalu	1
ksf	Bafia	0.777777778
lea	Lega	1
lin	Lingala (Kinshasa)	0.222222222
lol	Mongo-Nkunda	1
mcp	Makaa	1
ndg	Ndengereko	1
nso	Sotho, Northern	0.833333333
nya	Chichewa	0.944444445
sna	Shona	1
ssw	Swati	0.833333333
swh	Swahili	1
toi	Tonga	1
tsn	Tswana	0.777777778
ven	Venda	1
zul	Zulu	0.833333333

Table 8: Berber

ISO	Language	GCS
jbn	Nafusi	0.694444445
kab	Kabyle	0.694444445
shy	Tachawit	0.694444445
taq	Tamasheq	0.694444445
tzm	Tamazight	0.694444445
zen	Zenaga	0.694444445

Table 9: Chadic

ISO	Language	GCS
gid	Gidar	0.45
hau	Hausa	0.666666667
lln	Lele	0.466666666
mfk	Miya	0.6
pip	Pero	0.125
sur	Mwaghuvul	0.083333333

Table 10: Cushitic

ISO	Language	GCS
ahg	Qimant	0.533333333
awn	Awngi	0.611111111
bej	Beja	0.5
bsw	Baiso	0.416666667
dal	Dahalo	0.277777778
dsh	Daasanach	0.472222222
gax	Borana-Arsi-Guji Oromo	0.533333333
gdl	Dirasha	0.466666666
irk	Iraqw	0.433333334
ktb	Kambaata	0.361111112
rel	Rendille	0.533333333
som	Somali	0.5
tsb	Tsamai	0.444444445

Table 11: Eastern Nilotic

ISO	Language	GCS
kdj	Karamojong	0.722222222
mas	Masaai	0.5
tuv	Turkana	0.833333333

Table 12: Khoe-Kwadi

ISO	Language	GCS
hnh	ǁAni	0.533333333
naq	Nama	0.611111111
nhr	Naro	0.611111111
xuu	Khwe	0.466666666
kwz	Kwadi	0.25

Table 13: North-Central Atlantic

ISO	Language	GCS
bjg	Bydjogo	1
bqj	Bandial	1
cou	Wamey	0.833333333
ffm	Fulfulde	1
snf	Noon	0.777777778
srr	Serer-Sine (Serer)	1
wol	Wolof	0.777777778

Table 14: Semitic

ISO	Language	GCS
amh	Amharic	0.666666667
arb	Standard Arabic	0.611111112
ary	Moroccan Arabic	0.611111112
heb	Modern Hebrew	0.444444445
mlt	Maltese	0.777777778
tig	Tigré	0.611111112
trg	Lishán Didán	0.533333334

Table 15: Mel

ISO	Language	GCS
gol	Gola	0.611111111
tem	Timne	1
kss	Kissi	0.75

Table 16: Ta-Ne-Omotic

ISO	Language	GCS
bcq	Bench	0.611111111
kqy	Koorete	0.25
mdy	Male	0.555555555
wal	Wolaytta	0.555555555