A typology of northwestern Bantu gender systems

Di Garbo, Francesca

2022-07

Di Garbo , F & Verkerk , A 2022 , ' A typology of northwestern Bantu gender systems ' , Linguistics , vol. 60 , no. 4 , pp. 1169-1239 . https://doi.org/10.1515/ling-2020-0217

http://hdl.handle.net/10138/346441
https://doi.org/10.1515/ling-2020-0217

Francesca Di Garbo* and Annemarie Verkerk

A typology of northwestern Bantu gender systems

https://doi.org/10.1515/ling-2020-0217
Received September 24, 2020; accepted June 25, 2021; published online April 8, 2022

Abstract: Northwestern Bantu is the most linguistically diverse area of the Bantu-speaking world. Several unusual grammatical gender systems are reported for this area, but there has been a lack of comprehensive comparative studies. This article is a typological investigation of northwestern Bantu gender systems based on a sample of 179 languages. We study the distribution of various patterns of animacy-based agreement in the languages of the sample and in relationship with the Agreement Hierarchy. We find that animacy-based agreement is widespread in northwestern Bantu. If restricted to animate nouns, it tends to coexist in stable variation with syntactic agreement. When generalized to both animate and inanimate nouns, animacy-based agreement appears to contribute to the erosion of gender marking. In line with the prediction of the Agreement Hierarchy, we find that animacy-based agreement is prevalent with verbs and pronouns. Within the noun phrase, it spreads in ways that are suggestive of a hierarchy of syntactic integration between nouns and adnominal modifiers, which had gone unnoticed in the existing literature. These results have important implications for current models of Bantu gender systems and shed new light on animacy effects in the diachrony of gender more generally.

Keywords: agreement hierarchy; animacy distinctions; Bantu languages; grammatical gender; language evolution and change

1 Introduction

Grammatical gender systems, also known as noun class systems (Maho 1999; Katamba 2003), are one of the signature features of the Bantu language family, the

***Corresponding author: Francesca Di Garbo**, Department of Languages, General Linguistics, University of Helsinki, Helsinki, PL 24 (Unioninkatu 40) 00014, Finland,
E-mail: francesca.digarbo@helsinki.fi
Annemarie Verkerk, Department of Language Science and Technology, Saarland University, Saarbrücken, Germany, E-mail: annemarie.verkerk@uni-saarland.de

largest subgrouping within the Atlantic-Congo family.[1] Bantu gender systems tend to be represented as a monolithic block of complex nominal classification systems. They typically consist of more than five (often more than ten), non-sex-based gender distinctions,[2] and pervasive patterns of gender agreement, whereby the gender of nouns is indexed on a variety of agreement targets (typically, adnominal modifiers, verbs and pronouns). Gender assignment, that is, the tendencies and principles that regulate how nouns are allotted to a given gender (Corbett 2013c), is both semantic and formal. For instance, in Swahili (G41, swah1253),[3] names of plants are typically associated with gender 3/4, which is an instance of semantic gender assignment.[4] In addition, in Swahili, nouns that are marked by class 3/4 prefixes (*m-/mi-*) tend to trigger agreement in gender 3/4, independently of their meanings, which is an instance of formal gender assignment (Corbett 1991: 47–48). Finally, not all gender 3/4 nouns in Swahili denote plants and not all plant names are found in gender 3/4. This suggests that, in addition to being based on semantic and formal criteria, gender assignment rules in Swahili (as in many other Bantu languages) are also partially opaque.

Bantu gender systems are overall stable and have been reconstructed as already fully grammaticalized in Proto-Bantu (Schadeberg 2003: 149). They are generally well-described, both cross-linguistically and language-specifically, and often used as a model for the description of similar gender systems in other branches of the Atlantic-Congo family. In spite of this generally high degree of

1 We rely on Glottolog (Hammarström et al. 2019) for the genealogical classification of languages and language groupings discussed in this article. We are agnostic as to the arguments for or against the genealogical groupings assumed by Glottolog's classification, and use them as a mere reference system. This applies to our use of the term Atlantic-Congo. See also Nurse and Philippson (2003a) for a discussion of Bantu within Niger-Congo.

2 These categories are based on the classification of gender systems used by WALS, *The World Atlas of Language Structures* (Corbett 2013a, 2013b).

3 Here and in the rest of the article, the first mention of a language is accompanied by the respective Guthrie label and Glottocode (Hammarström et al. 2019). Guthrie (1948) classifies Bantu languages in zones named after alphabetical letters. Each zone is further divided into up to nine groups roughly consisting of sets of 10 languages each (for instance, zone A consists of group A10, A20, A30 etc.). Individual languages within each group are then referred to by non-round digits, such as A11 within A10, A21 within A20, and A31 within A30. The Guthrie's zones are a referential system based on similarities between languages that are geographically close to each other. This classification system does not aim to reflect genealogy. However, varying degrees of genealogical relatedness can be identified within the different zones. For instance, zone A and B are much less coherent from a genealogical point of view than zone S (Philippson and Grollemund 2019).

4 As detailed in Section 2, it is a tradition within Bantu studies to label genders after numbers. Most genders consist of pairings of singular and plural gender markers. The singular gender classes are labelled after odd numbers (e.g., gender class 3), while plural gender classes are named after even numbers (e.g., gender class 4).

stability, however, several Bantu languages with highly reduced gender systems (both in terms of number of gender distinctions and richness of agreement), or lacking gender altogether, are also attested across the family. Maho (1999) reports that Bantu languages with highly reduced gender systems typically have only two genders, the animate and inanimate gender, semantic gender assignment, and very little agreement. Moreover, in those languages where gender is entirely lost, number (singular vs. plural) may be the only grammatical distinction that is marked on the noun and/or through agreement. In some cases, agreement may be lost altogether.

In his comprehensive survey of Bantu gender systems, Maho (1999) states that Bantu languages where gender systems are reduced to the opposition between animate and inanimate nouns, or even lost, are only found in the western region of the northern Bantu borderlands. In these areas, Bantu languages are and have been historically surrounded by Ubangi (a distantly related Atlantic-Congo grouping), and Central Sudanic languages. The western northern Bantu borderlands are in turn part of a wider area, here referred to as northwestern Bantu (henceforth NWB), which is quite unanimously held to be the most linguistically diverse and divergent area of the Bantu-speaking world (see Nurse and Philippson 2003b: 165; Grollemund et al. 2018: 119, as well as Section 3.1). Several unusual grammatical gender systems are reported for this area (Maho 1999), but they have never been studied in depth, which is what we aim to do in this article.

One process of semantic and morphosyntactic reanalysis which occurs in several Bantu languages and may affect the relationship between gender assignment and gender agreement in ways that are possibly also relevant to understand the erosion of gender marking is what in Bantu-specific literature is known as "animate concord" (Maho 1999; Wald 1975). Animate concord occurs when patterns of gender agreement usually reserved for nouns assigned to gender 1/2, where the majority of human nouns are allocated, are (optionally) used with all animate nouns, irrespectively of their lexical gender. This is illustrated in Example 1, from Swahili. In this example, the animate noun for 'friend', lexically assigned to gender 9/10, triggers agreement in gender 9/10 on the possessive modifier, while the verb shows agreement in gender 1/2 agreement, that is the "human/animate gender".

(1). *Rafiki* *y-angu* *a-me-fika*
CL9.friend CL9-of.me CL1-PRF-arrive
'My friend has arrived'
(Wald 1975: 483) (Swahili, Bantu)

Animate concord is a type of semantic agreement, which we label *animacy-based agreement*, because it sets animate nouns apart from the inanimates based on the

agreement patterns that animate nouns (optionally) take.[5] Animacy-based agreement is described as a minor agreement pattern by Maho (1999: 129), as well as by Contini-Morava (2008: 162), who suggests that it may even be largely limited to the languages of coastal Kenya and Tanzania. Maho (1999: 140–142) further observes that, albeit rare, pervasive animacy-based agreement may play an important role in the diachronic evolution of Bantu gender systems and in the emergence of highly reduced systems of gender marking. Animacy effects in the evolution of gender systems are widely documented in the general typological literature, as detailed in Section 2.2, which provides additional empirical ground to this suggestion.

Understanding the distribution of highly reduced gender systems and animacy-based agreement in the larger NWB area, as well as their relationship to the more conservative systems with solely syntactic agreement that are also attested in this area is key to gain a clearer picture of Bantu nominal morphosyntax and of the historical processes that shaped it. Our focus is on the relationship between gender assignment (again, how nouns are allotted to a specific gender), and gender agreement (again, how the gender of a noun is indexed on a large variety of targets covering adnominal modifiers, predicates, and pronouns). This focus allows us to consider two main issues in greater detail than has ever been attempted before: (1) the frequency of animacy-based agreement and highly reduced systems of gender marking, and (2) their structural and semantic make-up. We also study whether the distribution of animacy-based agreement in NWB aligns with established typological generalizations, which are subsumed under the Agreement Hierarchy (Corbett 1979, 1991). This hierarchy predicts that patterns of semantic agreement (such as animacy-based agreement) are more likely to spread from agreement targets that are syntactically distant from the controller nouns, such as pronouns and predicates, to agreement targets that are linearly closer to nouns, such as adnominal modifiers, which is exactly what Example 1 shows for Swahili. The Agreement Hierarchy is thus a useful tool to interpret the varying agreement systems attested in NWB languages and their synchronic and diachronic relationships.

5 In this article, we use the label *animacy-based agreement* whenever a given agreement target (e.g., the independent personal pronouns or the attributive adjectives) marks animacy distinctions through its inflections. We use this label both in those cases in which animacy-based agreement alternates with syntactic agreement (as in Example 1), and in those cases in which all agreement targets available in a language categorically inflect on the basis of animacy contrasts. We refer to languages of the latter type as languages with 'animacy-based gender'.

In order to address these issues, we sampled 179 language solely coming from the NWB area.[6] This sampling procedure increases the chances of capturing as much diversity as possible, and enables us to better contextualize our findings in terms of the broader ecology of the area and the genealogical relationships that exist between the languages spoken in this area.

Our results show that animacy-based agreement is far more wide-spread in NWB than previously thought (Contini-Morava 2008; Maho 1999). They also confirm that highly eroded systems of gender marking are a minority in comparison with fully-fledged gender systems, which can feature marking of animacy distinctions in the form of optional semantic agreement. Yet, our sampling procedure allows us to capture more eroded systems than those identified by Maho (1999), and to more faithfully characterize their structural and semantic diversity. In addition, while the study aligns with the general predictions entailed by the Agreement Hierarchy, it also suggests that aspects of it, particularly pertaining to how different types of adnominal modifiers respond to the spreading of semantic agreement, are in need of refinement. These results have important implications on current models of Bantu gender systems and the typology and evolution of gender agreement more generally. They refine the picture of Bantu variation in the domain of grammatical gender, by bringing to the fore a more fine-grained characterization of animacy effects in the gender systems of the typologically most diverse area of the Bantu-speaking world (Grollemund et al. 2018: 119; Nurse and Philippson 2003b: 165). This endeavor results in a wealth of empirical data, which we use to formulate hypotheses about the insurgence and distribution of the various systems attested in the languages of the sample, and about animacy effects in the diachrony of gender systems more generally.

The article is structured as follows. Section 2 discusses previous literature on animacy effects in the functioning and spreading of semantic agreement and on typological variation in Bantu gender systems. Section 3 introduces the study design and data collection procedure, while Section 4 presents the results of the qualitative and quantitative analyses we conducted. Possible diachronic scenarios behind the distribution and evolution of NWB gender systems are presented in Section 5. A discussion of the findings and their implications for Bantu studies as well as for the typology and evolution of gender systems more generally follows in Section 6.

6 We investigated all NWB languages, but were not able to find enough information about the gender system of about 70 additional languages. See Section 3 for further details.

2 Background

2.1 Bantu gender systems

Gender marking in Bantu is always prefixal (Katamba 2003: 111), and, according to traditional descriptions of Bantu gender systems, one gender consists of combinations (or *pairings*) of singular and plural classes. In addition, some patterns of gender marking can be number invariant. It is a tradition within Bantu studies to use odd numbers to refer to singular classes and even numbers to refer to plural classes. Individual genders, that is, pairings of singular and plural classes, may be labeled after ordinal numbers.[7] Thus, gender I is the combination of class 1 (singular) and 2 (plural). Alternatively, the label 'gender 1/2' is also used. In this article, we use the term *class* to refer to unpaired singular and plural sets of markers, as, for instance, class 1 and class 2 with respect to gender 1. We use the term *gender* to refer to pairings of the singular and plural classes. An illustration of singular and plural patterns of gender marking in Swahili is given in (2), with examples from gender 1/2 ([2a] and [2b]) and 7/8 ([2c] and [2d]).

(2) a. *M-toto* *m-dogo* *a-mefika*.
CL1-child CL1-little CL1-arrived
'The little child arrived.'
b. *Wa-toto* *wa-dogo* *wa-mefika*.
CL2-child CL2-little CL2-arrived
'The little children arrived.'
c. *Ki-kapu* *ki-dogo* *ki-mefika*.
CL7-basket CL7-little CL7-arrived
'The little basket arrived.'
d. *Vi-kapu* *vi-dogo* *vi-mefika*.
CL8-basket CL8-little CL8-arrived
'The little baskets arrived.'
(example adapted from Katamba 2003: 111) (Swahili, Bantu)

As shown in the examples, the gender system of Swahili (and of many Bantu languages) consists of two sets of markers: the overt gender markers, which encode gender distinctions on nouns, and the agreement markers, which encode gender distinctions on a variety of agreement targets such as adnominal modifiers and

7 See Güldemann and Fiedler (2019) for alternative traditions in the labeling of the different genders.

verbs.[8] Common gender agreement targets across Bantu languages are: adnominal modifiers (e.g., adjectives, demonstratives, quantifiers, possessives), verbs, relative constructions, and pronouns of various kinds (personal pronouns, demonstratives). Patterns of gender marking on nouns and through agreement may or may not correspond to each other in terms of (a) the number of overtly coded distinctions, and (b) their formal realization. For instance, with respect to (a), in many Bantu languages, the nominal prefixes for class 1 and 3 are the same (*mu-*), but the two are systematically distinguished through agreement. Similarly, with respect to (b), in Swahili, the subject prefix for class 1 has a different formal realization (*a-*) than its nominal (and adjectival) counterpart (*mu-*) (cf. [2a]). In view of these recurrent mismatches, we argue that gender marking on nouns and through agreement in Bantu are best conceived of as two separate dimensions of analysis, both synchronically and diachronically (see Güldemann and Fiedler 2019 for a similar argument in the broader Niger-Congo context).

Bantu gender systems are non-sex-based (Corbett 2013b) and, as mentioned above, characterized by a combination of semantic, formal, and opaque criteria of gender assignment (Corbett 2013c). While the semantics of Bantu gender systems varies both within and across languages, Katamba (2003: 115–119) identifies a set of generalizations. Some classes can be defined in terms of animacy, with nouns denoting humans being typically assigned to gender 1/2, animal nouns often contained in gender 9/10, and plant and tree names in gender 3/4. Another relevant feature is size, as a certain number of classes (typically: 7, 8, 11, 12, 13, 20, 21) are associated with the encoding of diminutive and augmentative meanings. Marking of the infinitive (systematically associated with class 15) and various locational meanings (class 16, 17, 18) is also common but not present in all languages. Specialists disagree with respect to the general applicability of semantic criteria to the description of Bantu gender systems. For the purposes of this article, we assume that the prototypical Bantu gender system is non-sex-based and only partially biased towards the overt expression of animacy distinctions.

Patterns of gender assignment in Bantu languages may function as word formation strategies whereby nouns are derived from other nouns as well as from verbs. By manipulating gender assignment, speakers of Bantu languages can also modify aspects of the denotational semantics of nouns and/or the construal of the noun referent in a given discourse context.[9] When speakers change the gender of a

8 In many Bantu languages, overt gender markers may be preceded by a prefix, traditionally known as the *augment*, which varies according to the particular class the noun is assigned to. The functions of the augment vary within and across languages, but they tend to pertain to the encoding of definiteness, focus, and/or specificity. For an overview, see Katamba (2003).

9 Flexible gender assignment is, in fact, a distinctive feature of many African gender systems. For an overview of the phenomenon cf. Di Garbo and Agbetsoamedo (2018).

noun in order to encode diminutive, augmentative or locative meanings, the noun in question triggers agreement in these semantically motivated agreement classes rather than in its lexical gender. This is also an instance of semantic agreement. In this article, we are only concerned with instances of semantic agreement that are based on the encoding of basic animacy distinctions, what we call animacy-based agreement.[10] This phenomenon is introduced in Sections 2.2 and 2.3.

2.2 Animacy effects in the functioning and restructuring of gender systems

While the origins of the gender systems of several language families around the world are highly debated (see, for instance, Matasović 2004 and Luraghi 2011 on the origins of grammatical gender in Proto-Indo-European), patterns of restructuring in gender marking are well understood. One seemingly uncontroversial generalization regarding the evolution of gender systems is that animacy distinctions, i.e., the linguistic encoding of the distinction between (various types of) living and non-living entities, are likely to play a role in the semantic restructuring of gender systems. Such processes of restructuring foster the transition from (relatively) opaque to semantically predictable systems of gender assignment and gender agreement (Igartua and Santazilia 2018; Seifart 2018; Vihman et al. 2018). In the following, we put the Bantu-specific phenomena that we study in this article in a wider typological context and discuss documented animacy effects in the functioning of gender systems. Processes of animacy-based restructuring in gender systems are attested in other branches of the Atlantic-Congo family (see Faraclas 1986; Good 2012; Güldemann and Fiedler 2019; Marchese 1988; among others).

Animacy effects in the organization of gender systems are often linked to competing distributions of *semantic* and *syntactic agreement* with nouns whose referential semantics and formal gender assignment clash (Corbett 1991, 2006, as well as; Dahl 2000). A textbook example of alternation between semantic and syntactic agreement is the German noun *Mädchen*, 'girl/young woman', which denotes a female entity, but is grammatically neuter. In spontaneous discourse, speakers of German are likely to use both feminine and neuter patterns of gender marking in agreement with this noun, and the distribution of these competing

10 Animacy distinctions also play an important role in Bantu gender resolution rules, that is, in the processes affecting patterns of gender agreement with conjoined noun phrases. Gender resolution rules are not investigated in this article. For an overview, see Katamba (2003: 114) and Downing and Marten (2019: 283–285).

agreement patterns is determined by the type of target on which gender marking occurs. More specifically, feminine marking (in agreement with the referential semantics of the noun) is most likely to occur with personal pronouns, while neuter marking (in agreement with the lexical gender of the noun) is most likely to occur with attributive modifiers (Corbett 1991).

Nouns of the *Mädchen* type, which are fairly common in languages with grammatical gender, are labeled *hybrid nouns* in virtue of their mixed agreement preferences (Corbett 1991). The distributional tendencies attested in German with respect to the agreement patterns triggered by this type of nouns are also cross-linguistically robust and subsumed under a well-known implicational hierarchy, the Agreement Hierarchy (Corbett 1979, 1991, 2000), given in (3):

(3) *attributive > predicate > relative pronoun > personal pronoun*

The Agreement Hierarchy expresses the likelihood of semantic agreement to occur with different types of agreement targets across four different syntactic domains.[11] These are: the noun phrase (attributive), the clause (predicative), and the sentence/discourse (relative pronouns at the sentential level, and personal pronouns at the sentential/discourse level). The four syntactic domains represent different degrees of syntactic integration between the controller and the target. The hierarchy predicts that if a language has semantic agreement on attributive modifiers, it also has semantic agreement on predicates, relative pronouns and personal pronouns. Thus semantic agreement is most likely to occur in the agreement domains that are linearly more distant from the controller noun.

While patterns of semantic agreement may be used in parallel with syntactic agreement for long periods of time, the co-existence of semantic and syntactic agreement may also have long-term consequences on the evolution of gender systems. This tends to be the case when semantic agreement becomes generalized to all agreement targets and is triggered by a large class of semantically related nouns and not just by a handful of hybrid nouns (Corbett 1991: 248–259). For instance, in languages with non-sex based gender, nouns denoting animate beings but formally assigned to different genders may start triggering one and the same agreement pattern on personal pronouns. This agreement pattern may then be gradually generalized to all the other agreement targets until no trace of the former lexical genders of these nouns is left. This is what typically happens in Bantu

11 While the targets of agreement are the word classes that inflect according to a given morpho-syntactic feature (such as gender or number), the domains of agreement are the "syntactic configurations" in which controllers and targets may occur (Audring 2019: 17). For instance, in Italian (Indo-European, Romance), within the syntactic domain of adnominal modification, possible targets of gender and number agreement are the demonstratives, the definite and indefinite articles, and the attributive adjectives.

languages where semantic agreement with animate nouns, what we call animacy-based agreement, becomes obligatory with all agreement targets.

2.3 Animacy-based agreement in Bantu gender systems

About 24 noun classes are reconstructed for proto-Bantu but none of the presently spoken languages maintains them all (Katamba 2003: 103–105). While, in practice, this means that all Bantu gender systems are in one way or the other reduced as compared to the system reconstructed for the proto-language, not all instances of reduction are equally conspicuous, and various types and degrees of restructuring can be identified. In his comparative overview of Bantu gender systems, Maho (1999: 54) distinguishes two cutoff points, (1) languages with seven or more unpaired class distinctions, and (2) languages with up to three distinctions. Cutoff point (1) captures languages that retain many of the properties of a prototypical Bantu gender system, and that are therefore labelled by Maho as displaying a "traditional gender system". Cutoff point (2) identifies languages characterized by a "reduced or highly reduced system" whose functional underpinnings are most likely restricted to the marking of animacy and/or number.

Maho (1999) develops a typology of gender systems, which distinguishes between patterns of gender marking on nouns and via agreement, and classifies languages with respect to this bidimensional space of variation. For each of the two dimensions, five types are identified, ranging between traditional gender marking, various types of animacy-based marking, number-based marking, and no marking at all. An illustration of the typology is given in Table 1, with examples taken from Maho's and our own sample (the languages in boldface are some of the languages included in our data set, but not featured by Maho).[12]

Maho (1999)'s typology classifies Bantu gender systems on a scale from most conservative (i.e., traditional) to most innovative (i.e., reduced), where reduction is measured by looking at the role played by animacy and number distinctions on nouns and through agreement. A highly traditional system is one where animacy is only one of the organizational criteria that motivates patterns of lexical gender assignment, and number distinctions are systematically encoded cumulatively with gender. Highly innovative/reduced systems are those where (a) patterns of

12 Guthrie labels and Glottocodes of the languages mentioned in Table 1: Zulu: S42, zulu1248; Kinshasa Lingala: C30b, ling1263; Makanza Lingala (aka Northwestern Lingala): C30, nort3345; Yansi: B85, yans1239; Kituba, Congo: H10B, kitu1245; Kituba, DRC: H10A, kitu1246; Amba: D22, amba1263; Bera: D32, bera1259; Bila: D311, bila1255; Kako: A93, kako1242; Polri: A92, pomo1271; Komo: D23, komo1260; Mbati: C13, mbat1248; Pande: C12, pand1264; Bodo: D308, bodo1272; Homa: D304, homa1239.

Table 1: Bantu gender systems according to Maho (1999: 130–131). Gender marking through agreement: A = traditional; B = traditional + animacy-based; C = animacy-based + singular/plural; D = singular/plural; E = none. Gender marking on nouns: 1 = traditional; 2 = traditional + animacy-based; 2^{i} = traditional + plural; 3 = animacy-based + singular/plural; 4 = singular/plural; 5 = none.

	Nouns					
	1	2	2^{i}	3	4.	5
Agreement						
A	Zulu					
B	Swahili	Lunda				
C	M. Lingala		K. Lingala	Amba, **Bera** **Bila**, Kako	Mbati Pande	**Homa**
D	Yansi			**Polri**		
E	Kituba			**Komo**	**Bodo**	

gender marking (both nominal and non-nominal) are entirely devoted to the encoding of animacy distinctions, (b) former cumulative gender and number markers have been reinterpreted as singular and plural morphemes, or (c) no gender marking is left. In between these two extremes are languages which display animacy-based agreement. In these languages, the innovations mentioned above are confined to either (some) nouns (e.g., the animate nouns) or (portions of) the agreement system (e.g., verbs), and the marking of semantic agreement tends to be optional rather than obligatory (e.g., Type 1B in Table 1).

Three important generalizations are put forward in Maho's work. Firstly, only some of the logically possible types are attested. For instance, there are no languages where overt gender marking is completely animacy-based, and gender marking through agreement is traditional. This suggests that gender agreement patterns may be more sensitive to undergo restructuring than overt gender marking (the same was also found by Güldemann and Fiedler 2019 for Niger-Congo). Secondly, Maho (1999: 140) observes that while innovations affecting only gender agreement tend to occur in languages of wider communication, innovations that drastically affect both overt and agreement-based marking are typically attested in the northern areas of the Bantu-speaking world. Thirdly, Maho (1999: 123) states that animacy-based agreement tends to be under-reported in Bantu grammatical descriptions. This omission is likely to be a consequence of the fact that animacy-based agreement is stigmatized as an instance of incorrect or informal language use.

Wald (1975) is an earlier study of animacy-based agreement in Bantu gender systems. He studies 20 north-east Coastal Bantu languages and finds five types of

possible systems: (1) semantic agreement with animate nouns is obligatory all throughout the agreement system (Bondei, G24 bond1247); (2) semantic agreement with animate nouns is obligatory everywhere except with possessive modifiers (urban varieties of Swahili, illustrated in 1, see also Contini-Morava 2008); (3) semantic agreement with animate nouns is obligatory only outside the noun phrase (Kami, G36 kami1256); (4) semantic agreement with animate nouns is optional but preferred in all contexts (Chonyi, E72c chon1287); (5) semantic agreement with animate nouns is tolerated outside the noun phrase but rejected elsewhere (Sambaa, aka Shambala, G23 sham128).

This typology nicely aligns with the distribution of semantic agreement predicted by the Agreement Hierarchy (cf. Section 2.2): semantic agreement is more likely to occur (and first arise) on agreement targets outside the noun phrase than on adnominal modifiers. It also resonates with more recent observations by Van de Velde (2021), who argues that, within Bantu, different types of adnominal modifiers may show different degrees of sensitivity to the spreading of semantic agreement, with possessive modifiers being more resistant to agree semantically and more likely to agree syntactically.

What we further observe in a few languages from our sample, and constitutes a central finding of this study, is the fact that animacy-based agreement may sometimes also extend to inanimate nouns, in that all inanimate nouns come to trigger one and the same agreement pattern irrespectively of their lexical gender. When such development occurs, the gender system of a Bantu language may become even more fundamentally animacy-based, with a bipartite distinction between animate and inanimate nouns marked on some or all agreement targets. To the best of our knowledge, this phenomenon has so far gone unnoticed in comparative Bantu literature. In this article, we argue that this development is in fact crucial to understand the connection between the rampant expansion of animacy-based agreement and the reduction and erosion of gender distinctions, which has been suggested in previous literature (Maho 1999), but never really explained.

To sum up, previous studies on the typology of Bantu gender systems have identified a number of innovations related to the spreading of animacy-based agreement, which, in particular, sets animate nouns apart from the inanimates. This process is reminiscent of well-known tendencies in the typological literature on gender agreement. However, to date there has been no comprehensive study of these patterns, as Maho (1999) focused on the morphosyntax and reconstruction of Bantu gender systems, and Wald (1975) focused on a small sample of north-east Coastal Bantu languages. The present article aims to fill this gap.

3 Method

3.1 Northwestern Bantu

We collected data from five Guthrie zones, zones A, B, C, D, and H. The five target zones are illustrated in Figure 1. Zone A, B, C, D, and H are marked in five different shades of orange, whereas languages from the remaining Guthrie's zones, which are not featured in this study, are marked in black.

As Grollemund et al. (2018: 119) put it, the label NWB is generally used to refer to a geographical area where Bantu languages from zones A, B, C, and parts of zones D and H are spoken. In this sense then, our use of the term overlaps with a large part of the existing literature. However, it should be pointed out that alternative definitions of the NWB areas also exist. These tend to be narrower and to exclude zone D and H, and even parts of zones B and C (cf., for instance, Grollemund et al. 2015: 1397, Figure 1), or broader, bridging to languages of zone L and/or to the Bantoid groups, such as Mamfe, Grassfield or, more generally, South-Bantoid (for an overview of the literature and various definitions of NWB see

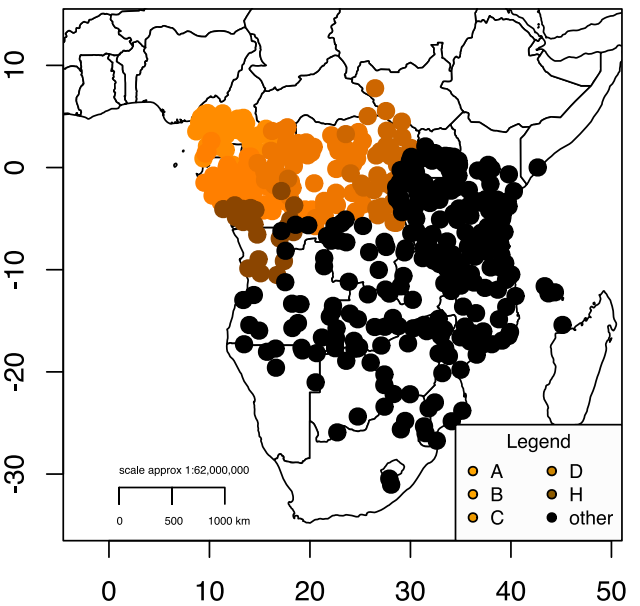

Figure 1: The Bantu languages according to Guthrie's zones. Languages belonging to zone A, B, C, D, and H, which our sample languages are selected from, are represented in different shades of orange. Black dots mark languages belonging to zones that are outside the area sampled for this study. The data points for this map are taken from Glottolog (Hammarström et al. 2019).

Grollemund et al. 2018: 119; see also Nurse and Philippson 2003b). Based on the analyses reported by Grollemund et al. (2015), the zones we chose to work with form a set of self-contained units that branch off the highest nodes of their proposed Bantu tree. This can be observed in Figure 12 in Appendix C: Additional visualizations. We mainly sample the North-Western (Cameroon and Gabon), Central-Western, and West-Western subgroups of the Bantu language family. Some of the Guthrie zone D languages we include are classified as Eastern Bantu languages in Grollemund et al. (2015), and some of the Guthrie zone H languages in our sample are classified as South-Western languages in Grollemund et al. (2015). However, in the Bantu tree by Koile et al. (Under review) which we use in Section 4.4 for visualization purposes and diachronic analyses, these D and H languages end up in different places in the tree, that is, closer to the Central Western languages (see Figure 8 in Section 4.4). In sum, as of today, NWB is not a strictly defined area, and the genealogical relationships of some languages on the border of Guthrie Zones D and H are not entirely clear. However, since we are not proposing a strict single diachronic analysis in this article, we argue that investigating these uncertainties is beyond the scope of this study.

3.2 Data collection

We started out by retrieving information about all Bantu languages from Glottolog (Hammarström et al. 2019), then added information on Guthrie's zones and limited data collection to zones A–B–C–D–H. In order to collect data on Bantu gender systems, we devised a questionnaire that aims at capturing their relevant structural properties at the language-specific level, and is also valid for crosslinguistic comparison. The questionnaire, which is given in Appendix A: Coding model, constitutes the basis of our variable design.

In order to investigate patterns of gender agreement in detail, and to pin down the interplay between the distribution of syntactic and animacy-based agreement, we gathered information about the patterns of gender marking exhibited by 14 different types of agreement targets, plus a category "other". These are: numerals, adjectival modifiers, possessive modifiers, demonstrative modifiers, quantifiers, question words,[13] verbs (including different types of grammatical relations), predicative adjectives, copulas, relative constructions (pronominal or other), personal pronouns, demonstrative pronouns, possessive pronouns, reflexives.

13 Our coding for question words encompasses both selective interrogatives that are used as adnominal modifiers, i.e., words for 'how many?/'which'? and interrogative pronouns such as 'who' and 'what?'. See Appendix A: Coding model for further discussion.

Finally, the category "other" is used in order to identify any additional host of gender marking which does not fall under those listed in the questionnaire. The fourteen targets were chosen in the attempt to cover for all four syntactic domains of the Agreement Hierarchy (personal pronouns, relative pronouns, predicative, and attributive) (Corbett 1979, 2000). We study the morphosyntactic behavior of individual agreement targets, rather than the agreement patterns associated with syntactic domains as a whole, because we are interested in unravelling possible differences between the inflections associated with different agreement targets within one and the same syntactic domain. Generalizations having scope at the level of agreement domains are formulated when applicable.

Short definitions of the comparative concepts we use to identify the different types of agreement targets, are given in Appendix A: Coding model. These definitions are a compromise between general typological and Bantu-specific literature. For instance, when defining the category "genitive/connectives", in addition to taking in to account general typological literature on adnominal possession, we also try to capture Bantu-specific patterns of encoding in this domain, i.e., the fact that connectives are used not only to mark adnominal possession, but also as a means to turn nominal property words into modifiers. Detailed illustrations of patterns of gender marking on a variety of target types are given in Section 4.1 (for instance, Example (4-b) illustrates subject agreement in Mokpwe, whereas Example (9-b) illustrates agreement with attributive adjectives in Ngelima).

For each of the chosen fourteen agreement targets, we code whether gender marking is syntactic, that is, based on the lexical gender of the noun (yes/no/no data), and/or semantic, that is, animacy-based (yes/no/no data). The coding design thus allows us to capture whether one, both, or neither type of marking is available for a given target type in a given language, or whether no information is retrievable from the sources.[14]

Our coding for animacy-based agreement aims at capturing whether any type of animacy distinction is marked on any of the fourteen target types, but does not differentiate between specific cutoff points along the Animacy Hierarchy (that is, whether the distinction is between "human" vs. "everything else" or "animate" vs. "inanimate") nor does it capture any constraints on the distribution of semantic agreement or whether animacy-based agreement may be overridden by other types of semantic agreement, such as diminutive or augmentative agreement. While we

14 Lack of syntactic and/or semantic marking on a specific agreement target and absence of a specific target type in the word class inventory of a language both result in a 'no' in our coding design. In other words, the design does not allow us to explicitly distinguish between languages where, say, reflexives are always gender-invariant and languages which lack the relevant lexical encoding of reflexivity altogether.

acknowledge this limitation, we consider our coding design to be accurate enough to provide a first comprehensive overview of the frequency distributions of syntactic versus animacy-based agreement in a large area of the Bantu-speaking world. This is something that had not been attempted in such a principled way before, if not for very narrow areas and a limited number of languages (Wald 1975). We provide more detailed qualitative analyses of the distinctions attested in some of the languages of the sample when we consider this to be crucial to understand how the different types of systems may have come to be and are related to each other (see Section 5).

The questionnaire design is such that one and the same language may be coded as displaying both syntactic and animacy-based agreement on one and the same agreement target. This is actually quite common in the languages of the sample and signals that the distribution of syntactic and semantic agreement within one and the same language is often not categorical, but rather subject to vary across speakers and usage contexts.

In addition, for each language of the sample, we collect data about (1) the number of singular, plural, and number-invariant class distinctions on nouns along with the number of singular/plural noun class pairings, and (2) the number of singular, plural, and number-invariant agreement classes, along with the number of singular/plural pairings of agreement classes. We include these counts in order to be able to compare the languages of the sample in terms of the overall number of overt gender markers as well as the number of agreement patterns. The questionnaire ends with 10 additional questions that aim to capture the obligatoriness of the attested agreement patterns, their interaction with number marking, as well as any additional data that would not fit elsewhere in the coding.

We collected information from grammars and other published or otherwise available materials on NWB languages, and tried to consult with specialists and/or native speakers where we found lacunae.[15] In total, we researched 255 languages, which are listed in Appendix B: The languages of the sample. Due to the poor state of language description in some areas, in this article, we present information on a sample of 179 languages. The dataset is included as Supplementary material.

15 Keith Beavon, Koen Bostoen, Ginger Boyd, Thera Crane, Muriel Garsou, Nadine Grimm, Myles Leitch, Jacky Maniacky, André Montingea Mangulu, Ruth Rahary, Christina Thornell, Lolke van der Veen, Mark Van de Velde, and Gert de Wit contributed their expertise on individual languages and/or language groupings, and helped us gathering materials about these languages. We thank them for their help.

3.3 Data analysis

We analyze our data through a combination of qualitative and quantitative methods. We begin by presenting a qualitative overview of the systems attested in the languages of the sample (Section 4.1).

We then develop a typology of animacy-based agreement in NWB by looking at the number of targets that receive syntactic versus animacy-based agreement (Section 4.2), as well as by using Multiple Correspondence Analysis (MCA, Appendix C: Additional visualizations). MCA is a method of data analysis which is very similar to the better known Principal Component Analysis (PCA). Both methods are used to detect and represent structures in a dataset by transforming potentially correlated variables into a smaller set of variables, called components or dimensions, which are no longer correlated and which best describe the variation attested in the dataset. While PCA is used for continuous variables and is thus not applicable to our dataset, MCA deals with categorical variables, like the ones we use in our questionnaire. MCA analyses were conducted using the package FactoMineR in R (Lê et al. 2008; R Core Team 2018).

We further move on to analyze the extent of syntactic and animacy-based agreement by presenting frequency distributions and correlation analyses of types of marking (syntactic vs. animacy-based) per agreement target (Section 4.3). The correlation analyses are done using the method developed by Pagel (1994) to model the evolution of two binary characters, which was later implemented by Revell in R (Revell 2012; R Core Team 2018). This method was first used in typology by Dunn et al. (2011) to analyze the evolution of pairings of word order features. In this article, we use it to test whether pairings of agreement targets (for instance, attributive adjectives and predicative adjectives, or numerals and quantifiers) show similar behavior both in the distribution of syntactic agreement and the distribution of animacy-based agreement. Frequency distributions and correlation analyses are used here to assess whether and how the distribution of syntactic and animacy-based agreement in the languages of our sample aligns with the predictions entailed by the Agreement Hierarchy.

In order to present a diachronic account of the proposed typology, we also use a combination of quantitative and qualitative analyses. Ancestral state estimation analysis is presented in Section 4.4. This is a phylogenetic comparative method that allows us to reconstruct what gender systems NWB languages may have had in the past, based on the current distribution of attested systems. These analyses were performed using the R package corHMM (Beaulieu et al. 2013; R Core Team 2018). The qualitative analyses, presented in Section 5, discuss selected synchronic distributions in the languages of the sample, which are suggestive of ongoing

patterns of variation and change and may constitute a connecting point between highly conservative and highly eroded gender systems.

For the correlation analyses and the ancestral state estimation analysis, we used a consensus tree from Koile et al. (Under review), which proposes a number of updates on the dataset by Grollemund et al. (2015), the latest phylogenetic analysis of the Bantu languages. The advantage of using the tree by Koile et al. (Under review) is that it includes more languages than Grollemund et al. (2015), with added languages being assigned to existing genealogical groupings on the basis of Glottolog (Hammarström et al. 2019). The result is a consensus tree that features all attested Bantu languages, at least as far as Glottolog's attestation records reach. In Figure 8, languages in Koile et al. (Under review)'s consensus tree that were not included in Grollemund et al. (2015) are prefixed by "Glotto", languages which *were* included in Grollemund et al. (2015) are prefixed by Guthrie code (these are named identically to how they appeared in Grollemund et al. 2015). Using Koile et al. (Under review)'s work enables us to include almost all sampled languages in the correlation and ancestral state estimation analyses, thus increasing their statistical power.

The code for these analyses and several Figures is included as Supplementary material.

4 Results

4.1 Qualitative overview of attested systems

This section showcases the diversity which we find attested in the languages of our sample with respect to patterns of gender marking and lack thereof. We represent the range of attested variation by first discussing instances of traditional gender systems and gradually moving on to systems which exhibit more or less pervasive animacy-based gender agreement or no gender agreement at all.

Mokpwe (A21, mokp1239), spoken in Cameroon, is an example of a language with a rather typical Bantu gender system. The language has seven singular and five plural overt gender markers, which result in nine singular/plural pairings. Likewise, there are seven singular and five plural agreement classes which, combined with each other, result in nine singular/plural pairings (Atindogbe 2013: 35). Mokpwe shows gender agreement in all expected syntactic domains and no animacy-based agreement is reported by our sources. Examples (4), (5) and (6) illustrate subject-verb gender agreement in Mokpwe with a human, animate and inanimate noun, respectively. Each of the examples provides illustrations both in the singular and the plural.

(4) a. *èmó-lánà* *à-lâ*
CL1-woman CL1-eat
'The woman eats.'
b. *βá-ălánà* *βá-lâ*
CL2-woman CL2-eat
'The women eat.'
(Atindogbe 2013: 55) (Mokpwe, Bantu)

(5) a. *é-lèlà* *é-ɲɔ̂* *má-léwá*
CL7-duck CL7-drink CL6-water
'The duck drinks water.'
b. *βé-lèlá* *βé-ɲɔ̂* *má-léwá*
CL8-duck CL8-drink CL6-water
'The ducks drink water.'
(Atindogbe 2013: 55) (Mokpwe, Bantu)

(6) a. *mó-ŏndó* *mó-óβì* *lì-βùmbú*
CL3-tail CL3-have CL5-hair
'The tail has hair.'
b. *mé-ŏndó* *mé-óβì* *mà-βùmbú*
CL4-tail CL4-have CL6-hair
'The tails have hair.[16]'
(Atindogbe 2013: 55) (Mokpwe, Bantu)

Eton (A71, eton1253), also spoken in Cameroon, is an example of another language with a fairly traditional system of gender marking. Van de Velde (2008: 290) points out one exception to this otherwise regular system. This is the singular form of the noun for 'chief', *ǹŋúŋúmá*, which is lexically assigned to gender 3/4, but systematically triggers class 1 agreement on the verb. This is shown in (7).

(7) *ǹ-kúŋúmá* *à-té* *kwàn*
3-chief 1-PR INF.BE.ILL
'The chief is ill'.[17]
(Van de Velde 2008: 290) (Eton, Bantu)

16 Noteworthy in this example is the fact that the noun for 'hair' is marked as plural when the subject ('tail') is plural. While this is an interesting illustration of the functioning of number marking and number agreement in the language, the specifics of this phenomenon are not discussed in our source and fall outside the scope of the investigation.

17 The present tense in Eton is a periphrastic construction consisting of the auxiliary verb *Lté*, which takes subject marking, and the infinitive of the lexical verb to be conjugated in the present (Van de Velde 2008: 254).

Conversely, with adnominal modifiers and when inflected as plural, agreement with this noun is always syntactic, that is in gender 3/4. Van de Velde (2008: 290) reasons that a tentative explanation for this anomaly resides in the fact that chiefs tend to be unique referents in a given discourse context, and unique reference is associated with gender 1/2 in Eton. Given that this is a very exceptional pattern in the language, which concerns only one noun, we consider Eton as an instance of a Bantu language with only syntactic agreement. We nevertheless think that pointing out this exception is a useful illustration of the fact that patterns of semantic agreement can intrude Bantu gender systems to varying degrees of pervasiveness, making it hard to break the diversity attested in the languages of the family into discrete types. This becomes more apparent as we move on to the analysis of systems where the presence of animacy-based agreement is more pervasive than in Eton, yet still largely optional.

A case in point is Lefa (A51, lefa1242), a NWB language spoken in Cameroon, which exhibits a traditional system of gender marking with instances of animacy-based agreement. In (8), the noun for 'chief' is a gender 5/6 noun and can trigger agreement on the verb either syntactically, that is based on its assignment to gender 5/6 as in (8a) or semantically, that is based on animacy, as in (8b) where the verb inflects according to class 1 subject prefix.

(8) a. *lɨ̀-fuəm* *ɗì-yúì*
CL5-chief CL5-came
'The chief came.'
b. *lɨ̀-fuəm* *á-yúì*
CL5-chief CL1-came
'The chief came.' (Isaac 2014: 9) (Lefa, Bantu)

As far as documented by our sources, animacy-based agreement in Lefa concerns only subject-verb agreement and is always optional.

A more pervasive instance of semantic agreement is one where the option of animacy-based marking extends to a higher number of agreement targets, if not all. This is, for instance, the case of Ngelima (C45, ngel1238), spoken in the Democratic Republic of the Congo. In his description of the gender system of the language, Gérard (1924: 17) reports that all animate nouns can take agreement in gender 1/2, irrespectively of their lexical gender, but that marking agreement based on lexical gender is also attested. Example (9), which shows alternation between syntactic and animacy-based agreement between a noun and its adnominal modifier, is used to illustrate this kind of alternation. The noun for 'crocodile' triggers class 3 agreement on the modifier in the singular (9a), based on its lexical gender, and class 2 agreement, based on animacy, in the plural (9b).

(9) a. *melanga* *m-endanda*
CL3.crocodile CL3-long
'long crocodile'
b. *melanga* *b-endanda*
CL4.crocodile CL2-crocodrile
'long crocodiles'
(Gérard 1924: 17) (Ngelima, Bantu)

Based on what we are able to infer from our sources, in Ngelima, the possibility to alternate between the two patterns is available on nearly all targets of gender agreement.

In our sample, we also find languages in which animacy-based agreement has come to be obligatorily marked at least on some agreement targets, most often the verbs, and with some nouns, most often animate nouns as opposed to inanimate. This is, for instance, the case of Ntomba (C35, ntom1248), spoken in the Democratic Republic of the Congo, where animate nouns lexically assigned to gender 9/10 obligatorily take agreement in gender 1/2 on the verb, as shown in (10), but otherwise retain their syntactic gender agreement on the other targets.

(10) *n-dzɔ́ɔ* *βá-lɛ́* *βicíndí*
CL10-serpents CL2-mordent/mangent talons
'Les serpents mordent/mangent les talons' ('Snakes bite/eat the heels', own translation)
(Motingea Mangulu 2010: 160) (Ntomba, Bantu)

A more extreme case of partially obligatory animacy-based agreement is attested in Lika (D201, lika1243), also spoken in the Democratic Republic of the Congo. At least in the domain of adnominal modification (adjectives, numerals, connectives, and demonstrative), Lika retains a fair amount of syntactic agreement.[18] Subject-verb agreement in Lika only distinguishes between animate and inanimate nouns. While animate subjects trigger the use of the verbal prefixes *a-/ø*, former class 1, in the singular and *ba-*, former class 2, in the plural, inanimate subjects can only trigger *a-/ø*, irrespectively of whether they are singular or plural (Augustin 2010: 18).[19] This is illustrated in the examples in (11), where the generalized animate and inanimate

18 A highly peculiar feature that characterizes gender marking in Lika is the fact that, for some of the noun classes, prefixal markers on nouns co-occur with noun class enclitics, which may have originated from one of the demonstrative paradigms (de Wit 2015: 201). This phenomenon is also attested in the neighboring languages Bwa (C44, bwaa1238) and Pagibete (C401, pagi1243) (de Wit 2015: 200). Some grammatical descriptions refer to the enclitics as suffixes.

19 As in many Bantu languages, many borrowings are assigned to class 1/2 in Lika. Thus class 1/2 agreement is also used with inanimate borrowed nouns (Augustin 2010: 18).

agreement markers are glossed as 3SG.AN/3PL.AN ([11a] and [11b]) and 3SG.INAN/3PL.INAN ([11c] and [11d]), respectively. Animate and inanimate agreement patterns are illustrated both in the singular and the plural.

(11) a. *mu-kó* *á-pʊng-á* *ndɪ* *ká-ĭnzɪnzíny-á*
1-woman 3SG.AN^{P}-start-FV^{P} P3 9b-REFL-complain-FV
'The woman started to complain'

b. *ɓombŭ* *ɓó-pik-og-o* *ɓa-ndáɓʊ* *na* *ɓe-nvunvú*
2-bird 3PL.AN-build-PL-FV 2-9.house with 2 + 9:9a-moss
'Birds build nests with moss.'[20]

c. *kó* *ngbíngó* *ɓé-motí* *áka,* *ɪ-ngbɔ́lɔ́*
PREP 1a.time 1NUM-one CT 9a-dogout
á-pung-a *kó-mw-óg-ó* *líɓó*
3SG.INAN-start-FV 9b-drink-PL-FV 5:water
'Suddenly, the dugout started to make water.'

d. *ma-ɗakĭ* *á-png-a* *kópúmúk-ó* *ɓí-kpŏ* *kpŏ*
6-pot 3PL.INAN-start-FV 9b-burst-FV MOD-kpŏ kpŏ
kpŏ
kpŏ
'The pots started to break "kpŏ kpŏ kpŏ"'.
de Wit (2015: 298, 299, 462, 283) (Lika, Bantu)

In all the examples discussed so far, animacy-based agreement, independently of its degree of obligatoriness and pervasiveness, only affects nouns denoting animate beings, while inanimate nouns continue to be distributed in several lexically-specified genders. Thus, for instance, in Ntomba, only a subclass of nouns (i.e., the animate nouns lexically assigned to gender 9/10) triggers obligatory animacy-based agreeement on the verb, while the rest continues to agree syntactically with their lexically-specified gender. In Lika, semantic agreement on the verb affects animate and inanimate nouns alike, and only two agreement patterns are available in this syntactic domain. One is almost exclusively used with animate nouns, and the other with the inanimates. In our sample, we find two more languages where animacy-based agreement has also extended to the domain of inanimate nouns. These are the closely related languages Mpiemo (A86c, mpie1238, Central African Republic, Cameroon and Congo) and the Bibaka variety of Ukhwejo (A802, ukhw1241, Central African Republic). Even though both languages retain instances of syntactic agreement, they both seem to be moving towards a system where the only type of distinctions that are flagged through agreement are animacy and number. This tendency is reported to be particularly

20 With class 9 nouns, nominal plurality is marked by the class 2 prefix.

prominent in the speech of the younger generations, while older speakers are more likely to use the traditional system.

Mpiemo has 11 distinguished singular/plural agreement patterns and 14 possible pairings of singular/plural noun classes (Thornell 2010). However, as further pointed out by Thornell (2010), what is noticeable at present in the speech of many speakers is that animate nouns systematically trigger agreement in gender 1/2, while the pairing 7/8 is systematically recruited for inanimate agreement (with inanimate nouns still keeping prefix 5 or 7 as their overt class marker in the singular and 6 or 8 in the plural). No detailed information is given by Thornell (2010) about the inflectional paradigm of specific agreement targets. A very similar pattern is attested in the Bibaka variety of Ukhwejo, referred to as Bendo in the work by Thornell (2012). Bibaka Ukhwejo retains eight different agreement patterns (four singular and four plurals) with five major and six minor possible singular/plural pairings. Besides these traditional, albeit already reduced, patterns of gender marking, gender distinctions are realigning around an opposition between animate and inanimate nouns. The pairing 1/2 is associated with animate nouns and the prefix *y-*, class 7, is used to mark agreement with inanimate nouns, irrespectively of their number. What is more, a tendency towards complete loss of gender distinctions is also noticeable, in that some speakers (particularly in the younger generations) generalize the use of agreement marker *y-* to animate nouns as well. In the work by Thornell (2012), variation between traditional and reduced, animacy-based, agreement is reported to run through the inflectional paradigm of possessive pronouns, demonstratives and the indefinite quantifier for 'some'. No information is given about other agreement targets. As mentioned in Section 2.3 and further argued in Section 5, we believe that, even though limited in number, languages like Lika, Mpiemo and Bibaka Ukhwejo are crucial to understand how highly eroded systems of gender marking may have evolved in this part of the Bantu-speaking world. More specifically, we argue that the ongoing variation observed across generations of speakers in two of these three languages, Mpiemo and Bibaka Ukwejo, offers a view into a possible diachronic pathway from soleley syntactic agreement and lexically-specified gender to solely animacy-based gender or no gender, which may possibly be applied to other languages too.

The cases discussed so far can all be described as displaying partial distributions of animacy-based agreement. We have seen that in those languages in which semantic agreement extends to a variety of agreement targets, as in Ngelima, it tends to remain in optional alternation with syntactic agreement. Conversely, if obligatory, it tends to be confined to selected agreement targets and, most typically, to patterns of subject-verb agreement, with only animate nouns agreeing semantically (as in Ntomba) or both animate and inanimate nouns (as in Lika). In such cases then, animacy-based agreement may be said to run in parallel

with syntactic agreement. As mentioned earlier on, however, in our sample, we also find instances of more pervasive restructuring of patterns of gender marking, where the entire agreement system revolves around the encoding of animacy contrasts. An overview of these systems of gender marking, and their main characteristics, is presented in the following, going from the least to the most radical instances of restructuring.

Nzadi (B85, nzad1234), spoken in the Democratic Republic of the Congo, retains four singular nominal prefixes, three plural prefixes, and six productive singular/plural pairings. The general pluralizer *ba-* can be optionally used to mark plurality, but not with nouns that have regular plural prefixes. Uncountable nouns are number-invariant. Patterns of gender agreement strongly diverge from the traditional Bantu type. While possessive constructions preserve relics of the traditional gender marking system,[21] the rest of the agreement system is organized around the opposition between human and non-human or singular and plural referents, as shown in Figure 2. Since at least the personal pronouns are based on the opposition between human and non-human referents, which is a type of animacy contrast, we classify Nzadi as displaying animacy-based agreement on this target type.

More pervasive instances of restructuring in patterns of gender marking are attested in those languages where both nominal and agreement marking deviate from the traditional Bantu gender type. This is, for instance, the case of Kako (A93, kako1242), spoken in the Central African Republic, where gender marking on nouns as well as through agreement are entirely based on the distinction between animate and inanimate nouns, with one dedicated agreement pattern each. This is

	singular		*plural*	
	human	*non-human*	*human*	*non-human*
pronouns **(3rd person)**	ndé	nɔ̌	bɔ̌	mɔ̌
demonstratives **(e.g. 'this')**	ná-pɛ		bá-pɛ	má-pɛ
adjectives **(some; e.g. 'big')**	o-nân		e-nân	
adjectives **(some; e.g. 'bad')**	o-bé			

Figure 2: Gender agreement in Nzadi. Data from Crane et al. (2011: 75).

21 In Nzadi, the genitive linker *é* is not used with nouns historically assigned to the Proto-Bantu singular classes 1 and 9, whereas it is used in the plural form of these nouns and with all other nouns (both in the singular and in the plural) (Crane et al. 2011: 78). Following the description by Crane et al. (2011), we do not consider this to be an instance of productive gender agreement.

illustrated in (12) through examples of noun-demonstrative agreement with an animate (12a) and an inanimate (12b) plural noun.

(12) a. *ɓè-Ngo* *ɓa-ka* *tì* *ɗɔkɔ̀* *na*.
AN.PL-cochon AN.PL-DEM NEG grandir NEG
'Ces cochons ne sont pas grands.' ('These pigs are not big', own translation)

b. *mɛ̀-kandɛ* *ma-ka* *ma* *lòlò*.
INAN.PL-habits INAN.PL-DEM déjà brûlé
'Ces habits sont brûlé.' ('These clothes are burnt', own translation)
(Ernst 1992: 36) (Kako, Bantu)

In Pande (C12, pand1264), also spoken in the Central African Republic, gender marking on nouns is entirely eroded, but the former class 2 prefix *ba-* — historically the plural of class 1, the Bantu 'human' class — is used as a general pluralizer for both animate and inanimate nouns. In contrast, Pande retains a productive system of subject-verb gender agreement (no other instances of agreement are retained), which is entirely based on the distinction between animate and inanimate nouns. This is illustrated in (13).

(13) a. *ŋgʉ́rɨ̀* *á-wà*
pig AN-will.die
'The pig will die.'

b. *ḿbʉ́lá* *ɛ́-hɨ́lá*
rain INAN-will.stop
'The rain will stop.'
(Richardson 1957: 35) (Pande, Bantu)

Somewhat the opposite situation is found in Polri (A92, pomo1271), spoken in Cameroon and Congo, where relics of the traditional gender system are left in the nominal domain whereas patterns of agreement on quantifiers and possessive pronouns only express number distinctions. More specifically, in the nominal domain, the singular prefix *mu-* and the plural prefix *bo-* are used to mark number distinctions with the human nouns for 'man', 'woman', 'spouse', 'person', 'child' (which historically belonged to class 1/2). With all other nouns, the singular is zero-marked and the plural is marked by the prefix *be-*. Examples of singular/plural number agreement on the quantifier for 'all' is given in (14).

(14) a. *ɓùtì* *ɓɛ̀-j^{w}ɔ̂*
PL.homme PL-IND
'tous les hommes' ('all the men', own translation)

b. *ɓɛ̀-nɔ̌n* *ɓɛ̀-jʷɔ̂*
PL-oiseau PL-IND
'tous les oiseaux' ('all the birds', own translation)
(Wega 2012: 128–129) (Polri, Bantu)

In our sample, we find five more languages in which nearly all traces of gender marking appear to be lost. This is the case of Bodo (D308, bodo1272), spoken in the Central African Republic, and Homa (D304, homa1239), spoken in Sudan, but nowadays nearly extinct. Bodo does not have any productive pattern of gender agreement apart from a human versus non-human distinction encoded on third person pronouns by means of the prefixes *yo-* and *ba-*, which are the singular and plural pronominal prefixes for human antecedents, and *-a* which is used for any other type of antecedent (Santandrea 1963: 94–95). The prefixes *mo-* and *bV-* in turn encode singular-plural distinctions on the noun (Santandrea 1963: 91), and are in all likelihood fossilized remnants of the Proto-Bantu noun class markers for class 1 and 2. In Homa, no traces of gender marking are left, except for a small class of adjectives, which are reported to take different inflections depending on whether the controller noun is human or non-human (Santandrea 1963: 96). This residual trace of animacy-based agreement is however only cursorily mentioned by our source.

An even more aberrant system, where animacy distinctions are not at the core of restructured gender marking, but only a condition on the distribution of plural marking, is what we find attested in Komo (D23, komo1260), spoken in the Democratic Republic of the Congo. According to Thomas (1994: 182), more than 200 nouns lack any inflectional prefix in Komo, while the prefix *ba-* can pluralize anything that is animate. According to Harries (1958: 269), some nouns can also take the plural prefix *i-*. These nouns are all inanimate. Neither traditional nor animacy-based prefixal agreement is left in the language. However, some form of reduplicative noun-adjective agreement has developed, whereby adjectives can be reduplicated when used attributively (Example 15). With inanimate nouns, adjectival reduplication only occurs in the plural (15b), whereas with animate nouns it occurs both in the singular and in the plural (15c and d).

(15) Reduplicative adjectival agreement in Komo
a. *endú* *ánje*
house red
'the red house'
b. *nkpá* *ánjenje*
person RED.red
'the red person'

c. *éndú* *ánjenje*
house RED.red
'the red houses'

d. *ba-kpá* *bá* *ánjenje*
PL-person being RED.red
'the red people'
(Thomas 1994: 193) (Komo, Bantu)

The phenomena attested in Komo do not align with any of the patterns which we encounter in languages with fully restructured but still productive, animacy-based gender systems, as for instance, Kako. Thus, Komo cannot be classified as a language with a productive gender system.[22]

Kituba and (Kinshasa) Lingala (Congo Kituba: H10B, kitu1245; DRC Kituba: H10A, kitu1246; Kinshasa Lingala: C30b, ling1263), the two Bantu creoles included in our sample, stand out for having the most peculiar make up of patterns of restructuring discussed so far. Both languages display highly eroded systems of gender agreement where agreement marking only encodes animacy and number distinctions (Kinshasa Lingala) or is completely absent (Kituba). Conversely, both languages display strikingly conservative patterns of class marking on nouns. In Kinshasa Lingala, we find seven distinct singular prefixes, five plural prefixes, three number-invariant prefixes and seven singular/plural pairings of nominal prefixes (Bokamba 1977; Meeuwis 2013). The variety of Kituba spoken in the Democratic Republic of Congo has six singular nominal prefixes, six plural, five-number invariant and six pairings of singular and plural noun prefixes (Mfoutou 2009; Mufwene 1997), while the Congo variety of Kituba has seven singular nominal prefixes, four plural, one number-invariant and seven pairings of singular and plural noun prefixes (Buchanan 1996/1997; Stucky 1978). Maho (1999: 140) argues that this type of development is typical for Bantu languages of wider communication, whereas restructuring affecting both noun-based and agreement-based marking tends to be restricted to the northern Bantu borderlands. While our data would seem to align with this observation, only a systematic survey of gender marking in Bantu languages of wider communication outside the northwestern area could confirm whether Maho's generalization also holds for the rest of the Bantu-speaking world.

22 In Komo, the distribution of plural marking on nouns and of reduplicative plural agreement on adnominal modifiers is constrained by the animacy of the noun. These types of plurality splits are in alignment with well-documented tendencies in the typological literature on number systems (Smith-Stark 1974).

Finally, in our sample, we also find one language whose highly reduced gender system does not seem to directly relate to any form of animacy-based agreement. This is Shiwa (A803, shiw1234), spoken in Gabon. The gender system of Shiwa is described by Ollomo Ella (2013) as displaying heavy restructuring in comparison with the more conservative systems attested in neighboring languages. In the singular, nouns can either be marked by their regular class or, alternatively, by a generalized class marker (whose nominal realization is zero or *N-*). Gender agreement complies with the gender marker carried by the noun, independently of whether this is marked by its regular class marker or by the generalized class marker. It is not clear from the source whether the alternation between the two types of overt gender marking is semantically motivated, but Ollomo Ella (2013: 203) identifies a clear pattern of generational shift, whereby younger speakers are more likely to use the generalized class marker than older speakers.

In this section, we have shown that a number of interrelated factors contribute to shape the distribution of the systems of gender marking attested in our sample. These factors ultimately point to two main dimensions of variation, i.e., whether animacy-based agreement is optionally and/or obligatorily available for all agreement targets or only for some of them, and whether animacy-based agreement applies only to animate nouns (with inanimate nouns triggering syntactic agreement with their lexical genders) or to both animate and inanimate nouns (with inanimate nouns also converging towards one and the same semantically-motivated agreement pattern). In the next sections, we continue to explore these matters with the help of quantitative methodologies.

4.2 A typology of gender marking in northwestern Bantu

In this section, we present a typology of NWB gender systems by looking at the gender inflections exhibited by each agreement target within and across languages.[23] In principle, each target type may be associated with one of the four logically possible configurations:

1. it may display only syntactic agreement
2. it may display only animacy-based agreement
3. it may display both syntactic and animacy-based agreement
4. it may lack both syntactic and animacy-based agreement

23 These analyses are not corrected for genealogical nor for geographic autocorrelation.

Given all logically possible combinations of the two agreement patterns attested in the languages of the sample (syntactic vs. animacy-based), we can posit four types, which are illustrated in the form of a tetrachoric table in Table 2.

As already illustrated in the qualitative overview presented in Section 4.1, these four logically possible types are indeed attested. Here we show that these four language types can also be distinguished bottom-up, that is, by aggregating the inflections associated with every agreement target across all languages of the sample. We argue that the two analyses, top-down and bottom-up, only partially overlap, which nicely illustrates the benefits of combining both approaches when searching for empirically grounded typological generalizations.

In Figure 3, we plot the languages of the sample by looking at how many agreement targets show syntactic agreement and how many exhibit animacy-based agreement. For the sake of coherence between the qualitative overview presented in Section 4.1 and the current section, some of the languages discussed in Section 4.1 are explicitly labeled in Figure 3.

In order to assess how the tetrachoric table relates to the number of targets that show syntactic versus animacy-based agreement, in Figure 3, we have color-coded the four groups listed in Table 2. Languages with only syntactic agreement cluster flat on the *x* axis of the plot (marked in black, 121 languages). Most commonly, they have between 6 and 12 different agreement targets. Languages with only animacy-based agreement, cluster closest to the *y* axis and very rarely display rich inventories of agreement targets (marked in orange, 11 languages). Languages with both syntactic and animacy-based agreement cluster towards the center-right area of the plot (marked in blue, 40 languages) and tend to have more targets agreeing syntactically than semantically. Finally, languages which score zero on both dimensions and thus lack any productive gender marking are located in the bottom left corner (marked in green, six languages).

The target counts show us a different way to conceive of the typological patterns posited through the tetrachoric table. First of all, we can clearly distinguish languages that display syntactic agreement from those that do not. Languages that

Table 2: Tetrachoric table of the logically possible types of agreement systems.

	Syntactic agreement	**Animacy-based agreement**
Type 1	True	False
Type 2	False	True
Type 3	True	True
Type 4	False	False

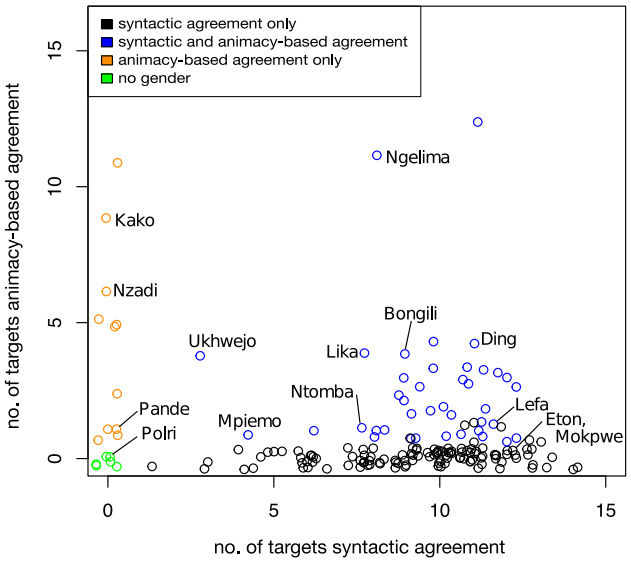

Figure 3: Number of targets that display syntactic agreement (*x*-axis) and animacy-based agreement (*y* axis). Points have been jittered so they do not overlap. Colors reflect the four-way typology introduced in the current section. Languages mentioned in Section 4.1 are labeled.

do not have syntactic agreement form a somewhat contiguous group stretching over a large area of the plot, with the number of targets that take animacy-based agreement ranging from zero (like Polri) to nine (Kako) or eleven (Bera). This cannot be said for languages with only syntactic agreement, whose distribution on the plot is somewhat more clustered. These languages mostly mark gender on six to twelve different targets, while very few such languages display gender agreement on less than five targets. Put another way, the (genealogically uncorrected) mean number of targets that inflect for gender in languages with only syntactic agreement is 9.3, SD = 2.6. While this is perhaps a given for Bantuists, who appreciate the fact that in traditional Bantu gender systems, gender agreement is pervasive, it is certainly notable from a statistical point of view. The pattern is also shared with the languages that mark both syntactic and animacy-based agreement, where the mean number of targets receiving syntactic agreement is 9.7, SD = 2.0. In these language, the number of targets that receive animacy-based agreement is clearly centered between one and five targets, with a mean of 2.6 targets, SD = 2.3.

In addition, while languages displaying both syntactic and animacy-based agreement could have been scattered all over the plot space, we find that this is not the case. What we observe instead is that there are no languages where animacy-

based agreement is possible for a greater number of agreement targets than those allowing for syntactic agreement. The only exception to this pattern is Ngelima, where syntactic and animacy-based agreement are possible on all agreement targets but verbs, which, based on what we infer from examples provided in our source (Gérard 1924), only display subject marking if the subject is animate. In addition, we find only very few languages that mark an approximately equal number of targets for both syntactic and animacy-based agreement. These are, for instance, Bibaka Ukhwejo and Mpiemo, which, as discussed in Section 4.1, are currently seemingly shifting towards a fully animacy-based gender system.

The 51 languages characterized by the co-presence of syntactic and animacy-based agreement or by solely animacy-based agreement are represented in Figure 4, where we show the distribution of both agreement patterns across languages and target types. The languages with the most eroded systems of gender marking (i.e., no syntactic agreement) are placed towards the bottom end of the figure (i.e., from Lingala to Bodo).

Figure 4 nicely matches the pattern anticipated above in that it shows that in languages with both syntactic and animacy-based agreement, the latter is never more pervasive than the former in its distribution across target types. In addition, the figure shows that the target types that are most frequently associated with animacy-based agreement across the languages of the sample are verbs and personal pronouns. While few languages have both syntactic and animacy-based agreement running across extensive parts of their agreement system (e.g., Ngombe and Ngelima), most languages display this possibility only on verbs and personal pronouns (e.g., Bangi, C32, bang1354), and a few others also in the adnominal domain (e.g., Ukhwejo or Ligenza, C414, lige1238). In Section 4.3, we discuss these patterns in light of the Agreement Hierarchy, while a diachronic intepretation of the data is proposed in Section 5.

To conclude, using bottom-up approaches to analyze the sampled data as we did in this section means that we can go beyond the patterns suggested by the tetrachoric table given in Table 2. More specifically, these approaches allow us to capture how the four discrete types posited in the table interact with each other, and to construct a more fine-grained picture of the typological profiles of gender marking in NWB. Once again, the data support a four-way classification of types of language structures, which mutually interact with each other in the ways highlighted in this section: languages with only syntactic agreement, languages with both syntactic and animacy-based agreement, languages with only animacy-based gender, and languages with no gender at all. This four-way classification can also be observed in the Multiple Correspondence Analysis (MCA) reported on in Appendix C: Additional visualizations.

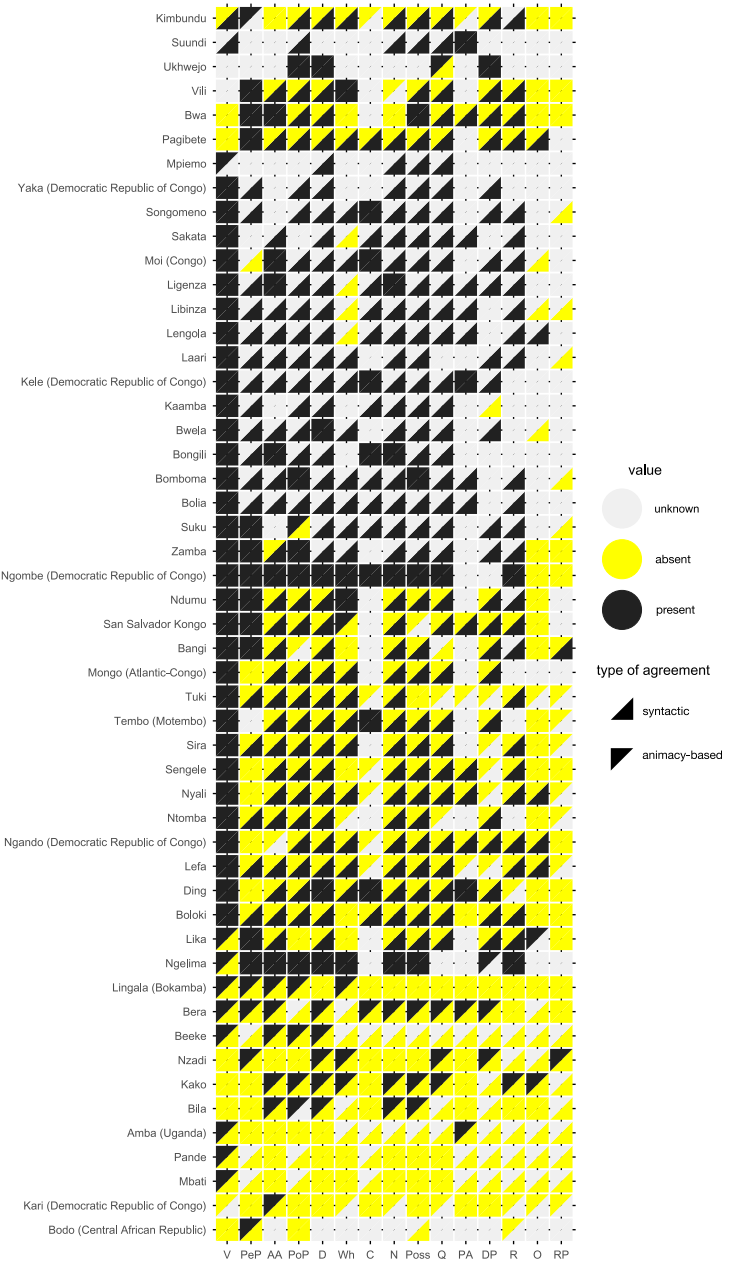

Figure 4: Distribution of syntactic and animacy-based agreement per language and across target types. V = verbs; PeP = personal pronouns; AA = attributive modifiers; PoP = possessive pronouns; D = demonstrative modifiers; Wh = question words; C = copulas; N = numerals; Poss = adnominal possession; Q = quantifiers; PA = predicative adjectives; DP = demonstrative pronouns; R = reflexives; O = other; RP = relative pronouns and other relative constructions.

4.3 Distributional analyses in light of the Agreement Hierarchy

In this section, we start with a simple distributional overview of how often each target receives a certain type of gender agreement, syntactic versus animacy-based (Figure 5). Note that this overview is not corrected for genealogical or spatial autocorrelation and serves only to show aggregate distributions. We go beyond this overview by conducting correlation analyses between the behavior of individual agreement targets. These analyses are controlled for genealogy and are presented in Figure 6. The patterns presented in this section are analyzed in light of the Agreement Hierarchy (Corbett 1979, 1991, 2000): attributive > predicate > relative pronoun > personal pronoun. The hierarchy predicts that the likelihood of semantic agreement, in this context animacy-based agreement, is highest in the domain of personal pronouns and lowest in the domain of adnominal modification.

Figure 5 represents the distribution of types of gender agreement, syntactic versus animacy-based, on the different types of agreement targets we coded for. The plot on the left hand side of the figure represents the frequency of occurrence of syntactic agreement per target type, whereas the plot on the right hand side illustrates the frequency of occurrence of animacy-based agreement across the same target types. The ordering of agreement targets within each of the two graphs is based on how often a given type of agreement is present on a given target type and thus differs across graphs. See Appendix C: Additional visualizations for an additional Figure, where the order is based on ratio of present-absent. The figure distinguishes between three levels of coding: presence (black), absence (white), and unknown (gray).

Figure 5 shows that syntactic agreement is overall much more frequent than animacy-based agreement. In addition, syntactic agreement is most common with (at least some types of) adnominal modifiers: 91% of the languages which we have data on have syntactic agreement on demonstratives. Adnominal modifiers are followed by verbs (85% of languages), relative pronouns (82%), and pronouns (77%). This nicely matches the generalizations entailed by the Agreement Hierarchy, whereby syntactic agreement is most likely to occur on adnominal modifiers, followed by predicative expressions and relative pronouns, with the personal pronouns being the least likely to have syntactic agreement.

As hinted at by the examples illustrated in Section 4.1 and Figure 4, animacy-based agreement (to the right of Figure 5) is most common on verbs (33%) and personal pronouns (15%). This is at least partially in line with the Agreement Hierachy, which lists the predicative and personal pronouns' domains as the most frequent attractors of semantic agreement, but in the reverse order (personal

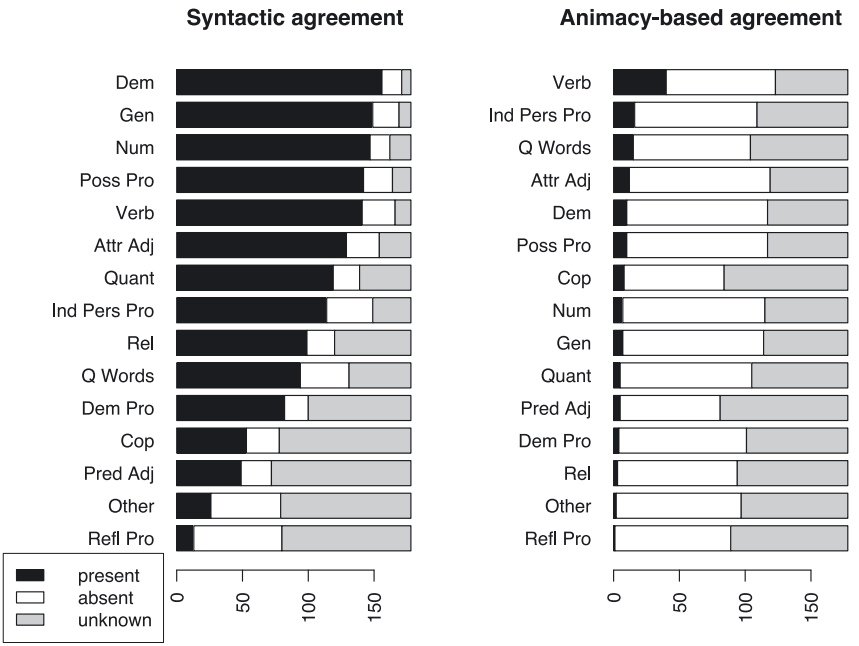

Figure 5: Distribution of syntactic and animacy-based agreement for all targets.

pronouns followed by predicates). The fact that, in our data, verbs override pronouns in being the strongest attractor of animacy-based agreement could be linked to the very nature of argument marking in Bantu languages, which has a chiefly anaphoric function (Bearth 2003: 122): while subject marking on the verb by means of gender prefixes does not require the presence of an overt nominal or pronominal subject, the opposite (overt lexical subject without number marking on the verb) is ungrammatical. In this sense then, the higher frequency of verbs as preferred locus for animacy-based agreement in comparison with pronouns would not contradict the crosslinguistic tendencies captured by the Agreement Hierarchy, but could be framed as a Bantu-specific construction which fully aligns with them.[24]

24 Initially, we coded separately for the marking of subject, object, and indirect object on verbs. After reviewing the data, we decided to collapse this distinction, because it was often hard to find information on object marking, especially indirect object marking, and because we found no languages that marked syntactic agreement only on objects but not on subjects. In the case of animacy-based agreement, interestingly, we did find languages in which gender marking on objects diverged from subject marking. According to our sources, Nyali (D33, nyal1250), San Salvador Kongo (H16a, sans1272), and Tuki (A601, tuki1240) allow animacy-based marking of

A few more patterns can be inferred from Figure 5, which partially depart or add upon the predictions entailed by the hierarchy. Both graphs of Figure 5 reveal that different types of adnominal modifiers may exhibit different degrees of propensity towards one or the other type of marking (syntactic vs. animacy-based). Notably, demonstratives (in 91% of the languages), adnominal possessors (88%), numerals (91%) and possessive pronouns (87%) are more frequently associated with syntactic agreement than attributive adjectives (84%). This could be, once again, a by-product of family-specific characteristics. For instance, several Bantu languages lack dedicated lexical classes of attributive adjectives or, if they have them, these can be gender-invariant. A more general explanation, which we put forward as a speculative thought in need of further empirical testing, is that the different agreement preferences shown by different types of adnominal modifiers reflect varying degrees of syntactic integration between nouns and their modifiers within a noun phrase.[25] Demonstratives, adnominal and pronominal possessors as well as numerals may have stronger syntactic ties with nouns than adjectives and quantifiers, and this would be reflected by their stronger sympathy for syntactic agreement. In turn, in the noun phrase, animacy-based agreement is more likely to appear on adjectives (occurs in 10% of the languages) than on demonstratives (9%), possessors (6%) and numerals (6%), as shown by the right hand side graph of Figure 5.

These patterns also match recent observations by Van de Velde (2021), who connects the distribution of animacy-based agreement within the noun phrase to what he calls the "Adnominal Modifier Apposition and Reintegration" mechanism (or AMAR) in Bantu languages. AMAR is the process whereby, in many Bantu languages, adnominal modifiers tend to be nominalized, apposed, and eventually syntactically and/or prosodically reintegrated to the noun phrase in which the modified noun occurs. This results in structures of the type 'the big men' versus 'the men, the big ones' (Van de Velde 2021: 6), which typically carry a contrastive function and thus contribute to facilitate reference identification. Interestingly, in Bantu languages with pervasive animacy-based agreement, this tends to apply to

objects on verbs, but not of subjects. Since this is only a marginal pattern, we merged the dataset so that marking of gender on the verb means either marking of the subject or the object, or both.

25 The existence of hierarchical effects which regulate the spreading of semantic agreement within the noun phrase has been sparsely noted in the literature, but never studied through systematic crosslinguistic comparisons. Cf. the work by Karatsareas (2009, 2014) on semantic agreement and gender loss in the Asian Minor Greek dialects, as well as Van Epps (2019) on gender loss in Jamtlandic, a Scandinavian variety spoken in Sweden, at the border with Norway. In both cases, definite articles, which are linearly closest to nouns, are the only agreement target where the traditional, fully fledged system of gender distinctions survives, while this appears to be lost, or highly eroded, on other agreement targets.

all adnominal modifiers but the possessives, which are also typically excluded from undergoing the AMAR mechanism, possibly due to their inherently selective semantics (Van de Velde 2021: 13). According to Van de Velde, the connection between the AMAR mechanism and animacy-based agreement may reside in the fact that when AMAR occurs, and modifiers are apposed to the noun phrase, they become linearly more distant from nouns and thus more sensitive to animacy-based agreement.[26]

The scenario evoked in Van de Velde's work closely matches the distribution of types of agreement per adnominal modifier that we observe in our data. As shown by Figure 5, more inherently selective adnominal modifiers, such as possessives, demonstratives, and numerals are less likely to show animacy-based agreement than, say, adjectives. While these tendencies offer promising evidence in support of the existence of a hierarchy of syntactic integration between different types of adnominal modifiers within the noun phrase, which manifests itself through the distribution of animacy-based agreement and possibly also the AMAR mechanism, this can only be confirmed through systematic empirical investigations of the syntax of Bantu noun phrases, which goes beyond the scope of the present study.

Besides these clear cut patterns, which nicely match with what we expect based on previous literature on the distribution of syntactic and semantic agreement, a few quirks in the distribution of preferred types of agreement across targets remain. We suggest that at least some of these quirks may be explained as a function of our coding design. For instance, Figure 5 shows that question words pattern closer to pronouns than adnominal modifiers, with respect to both syntactic and animacy-based agreement. Most likely, this is related to the fact that our coding for this target type encompasses both adnominal and pronominal question words, and that interrogatives pronouns for 'who?' and 'what?' tend to consistently encode basic animacy contrasts rather than lexical gender.[27] These are also more systematically described by our sources than other types of interrogatives. For a number of agreement targets, unclear or inconsistent distributional patterns may

26 Unusual word order patterns in Bantu noun phrases, which seemingly contradicts the generalizations entailed by Greenberg's Universal 20 (Greenberg 1963), can also be explained by virtue of the AMAR mechanism. For a discussion, see Van de Velde (2021) and references therein.

27 If a language in our sample displays animacy-based marking only on question words, we do not consider this to be an instance of animacy-based agreement. This is because, as detailed in Appendix A: Coding model, animacy-contrasts in the domain of question words usually occur on non-selective interrogatives, which cannot be described as agreement targets in the proper sense of the term (Idiatov 2007). Languages that only display animacy-based marking on question words are thus classified as only having syntactic agreement.

result from lack of data. This may be the case for predicative adjectives, copulas, and the category "other". Finally, the results for the reflexive pronouns seem to match with the fact that within Bantu, these are often gender-invariant prefixes attached to the verb stem, or independent words that may agree with the lexical gender of the noun via the use of pronominal prefixes.

In addition to examining which agreement targets are more often associated with syntactic and/or animacy-based agreement, we also test which targets behave similarly with respect to the agreement patterns they tend to be associated with. We do this by testing genealogy-informed correlations between, on the one hand, each pair of targets for their patterns of syntactic agreement, and, on the other hand, each pair of targets for their patterns of animacy-based agreement.[28] The results are summarized by the heatmaps in Figure 6, where the colors capture the p-values of the pairwise correlations, with blue representing significant correlations (with p-values <0.00024, see legend on the right and footnote 28).

Almost all targets are highly intercorrelated in their patterns of syntactic agreement, except for four: predicative adjectives, question words, reflexive pronouns and other. The low interconnectedness of question words may be due to the fact that both interrogative modifiers and pronouns are captured by this category. Low correlation levels for predicative adjectives, reflexive pronouns and other are probably due to the fact that gender marking on these targets is less common or rare. As can also be observed in Figure 5, predicative adjectives, reflexive pronouns and the category "other" are the three targets where we found the least amount of syntactic agreement, and for over half of the languages it is unclear whether they actually have any type of gender marking on those targets. Aside from these four targets, all other targets are highly intercorrelated, which implies that overall, languages are likely to either have/or not have syntactic agreement with them.

A possible outcome that we do not observe in the top plot of Figure 6 are opposing groups of targets correlating with each other, for instance a split between

28 We performed genealogy-informed correlation analyses for each possible pair of agreement targets, which means that we conducted $(15 \times 15) - 15 = 210$ different analyses twice, once for syntactic agreement and once for animacy-based agreement. As this means that we used a single dataset to run a large amount of tests, with each agreement target being tested 15 times for each of the two analyses, we applied Bonferroni correction to the results. Given that the desired significance level is $p = 0.05$, we divided this by 210 to get a corrected significance level of 0.00024. We did not consider correlations across type of marking, i.e., the correlation between syntactic and semantic agreement for attributive adjectives. Note that each analysis was performed on the full dataset (179 languages) while excluding languages where information on one of the targets involved in the pairwise correlation was missing.

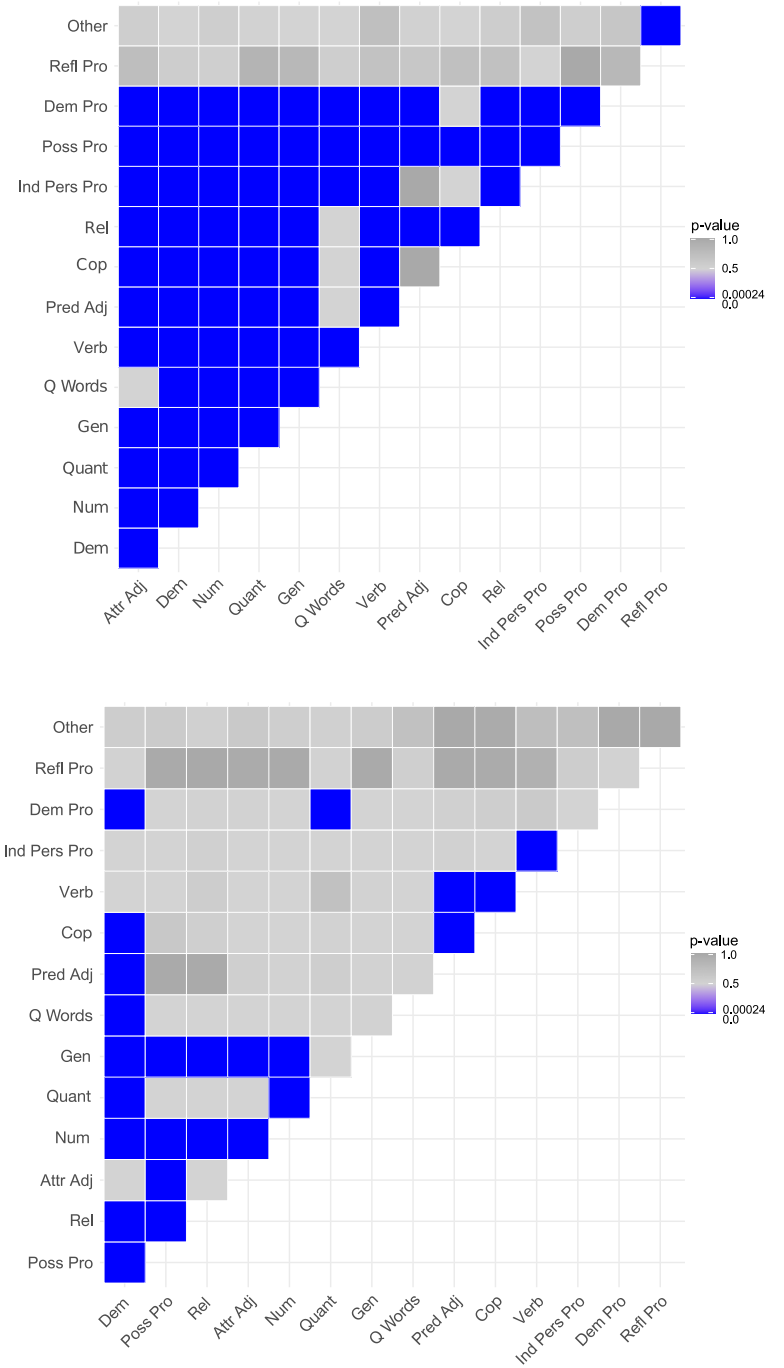

Figure 6: Heatmaps of *p*-values of correlation tests between each pair of agreement targets for syntactic agreement (top) and animacy-based agreement (bottom).

a group of adnominal targets that correlate with each other, and a group of predicative or pronominal targets that correlate with each other. Such a pattern could be suggestive of functional differences between different agreement targets. This is rather what emerges from the correlation analyses in the domain of animacy-based agreement, the bottom plot. A group of highly correlated targets can be identified in the left-most bottom corner of the heatmap, which rather neatly captures the domain of adnominal modification: attributive adjectives, demonstrative modifiers, demonstrative pronouns, genitives, numerals, possessive pronouns, quantifiers, and relative pronouns are intercorrelated. Towards the central-upper part of the plot, animacy-based marking on independent person pronouns is highly intercorrelated with animacy-based marking on verbs, while verbs are in turn highly intercorrelated with copulas and predicative adjectives. Note that copulas and predicatives adjectives are only further correlated with demonstrative modifiers, and not with any other adnominal targets. Independent personal pronouns are not correlated with any adnominal target, suggesting that these domains operate independently in attracting animacy-based agreement. These groupings match well-known functional differences between domains of agreement and are also in line with the tendencies unveiled by the right-hand plot on Figure 5. They ultimately confirm that presence of animacy-based agreement on independent personal pronouns is likely to go hand in hand with animacy-based marking in the predicative domain, and that adnominal modifiers also harmonize with each other in exhibiting, or lacking, animacy-based agreement.

That analogous grouping effects do not emerge so clearly from the distribution of syntactic agreement in Figure 6 might be explained by the fact that this type of marking remains highly pervasive among the languages of the sample. Thus, while the distribution of syntactic agreement does not differ much across types of targets and agreement domains, the distribution of animacy-based agreement is more target- and domain-specific.

4.4 A genealogical and geographical view on animacy-based restructuring

We have shown that animacy-based agreement is widespread in NWB, affecting, in one form or another, one third of the languages we found data on. In this section, we show how syntactic and animacy-based agreement are distributed in terms of

geography and genealogy. Figure 7 presents the distribution of gender systems in terms of the four-way typology presented in Section 4.2. What we can observe is that languages of the same type cluster together. In Cameroon, Gabon, and Congo-Brazzaville, languages with only syntactic agreement prevail. Towards the east and south, in the Democratic Republic of the Congo and the north of Angola, we find gender systems with both syntactic and animacy-based agreement. Languages with only animacy-based gender or no gender at all are even more clustered. They are found in the north and east on the border of the Bantu-speaking area, as well as in the southwest of the Democratic Republic of the Congo. Since some polygons are small and it is difficult to see their type, we include a point-based map in Appendix C: Additional visualizations.

The geographical distribution of languages with only animacy-based gender or no gender at all might be shaped by contact with existing and extinct non-Bantu languages. Ubangi and Central Sudanic languages spoken to the north of the NWB-speaking area either do not have gender systems or their gender systems are divergent from typical Bantu systems in that they involve sex-based or animacy-based distinctions (Boyd 1989; Corbett 1991; Dimmendaal 2000).[29] Before the arrival of Bantu, Ubangi, and Central Sudanic speakers, the Central African rainforest was inhabited by native populations commonly known as "Pygmies", who still reside in these areas. Recent work (Bostoen and Gunnink forthcoming) proposes that atypical features displayed by Bantu languages spoken in the Central African rainforest might be substrate effects from the native languages of the Pygmies, which are no longer spoken today. The highly restructured and eroded gender systems that we found in the north and east of the Democratic Republic of the Congo might be one of these atypical features.

We next turn to Figure 8 that displays the typological classification we presented in Section 4.2 on the tips of a phylogenetic tree, combined with reconstructions of gender system type on the internal nodes of the tree. For

29 See also the recent contribution by Fiedler et al. (2021) on the gender system of the Ubangi language Mba (mfc), spoken in the Democratic Republic of the Congo. The gender system of Mba is organized around two main patterns of agreement: syntactic agreement, whereby agreement patterns are largely determined by the different classes nouns are formally allocated to, and animacy-based agreement, which is entirely predictable on semantic grounds, based on the opposition between masculine humans and other animates in the singular (this distinction is neutralized in the plural where the same anaphoric pronoun is used for all plural animates). While the former system operates in the domain of adnominal modification, animacy-based agreement is restricted to the third person pronouns. This distribution is reminiscent of what we observe in those languages of our sample with both syntactic and animacy-based agreeement, with the main difference being that, in Bantu, animacy-based agreement never encompasses sex-based distinctions.

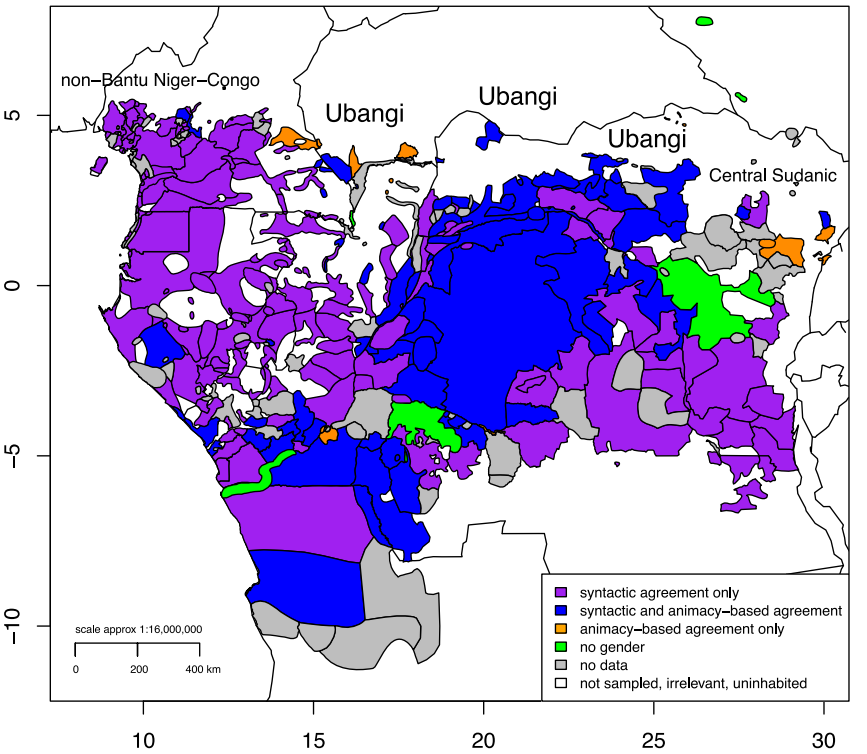

Figure 7: Distribution of types across the sampled area. Language polygons are taken from the world Language mapping System (Global mapping International, 2015. World language mapping system, version 17. Colorado Springs, CO (http://worldgeodatasets.com)) and constructed on the basis of language materials. White areas represent uninhabited land or bodies of water, towards the north also other non-Bantu languages not drawn, towards the east and south, other Bantu languages not sampled.

readability purposes, and because the higher-order subgrouping of Koile et al. (Under review)'s tree does not concern us here, only the part of the tree that features NWB languages has been plotted. In Figure 8, the major clades of the NWB languages have been labeled using the labels from Grollemund et al. (2015) (see Figure 12 in Appendix C: Additional visualizations for how these clades are positioned within the Bantu family as a whole). The ancestral state estimation analysis was constrained so that the ancestor of all sampled languages, Proto-Bantu, had a gender system with only syntactic agreement, which helps to make the ancestral reconstruction of "no data" become less prevalent. We did not want to exclude languages with no data from this figure, as this may have lead to a skewed understanding of the distributions. Proto-Bantu certainly had syntactic agreement

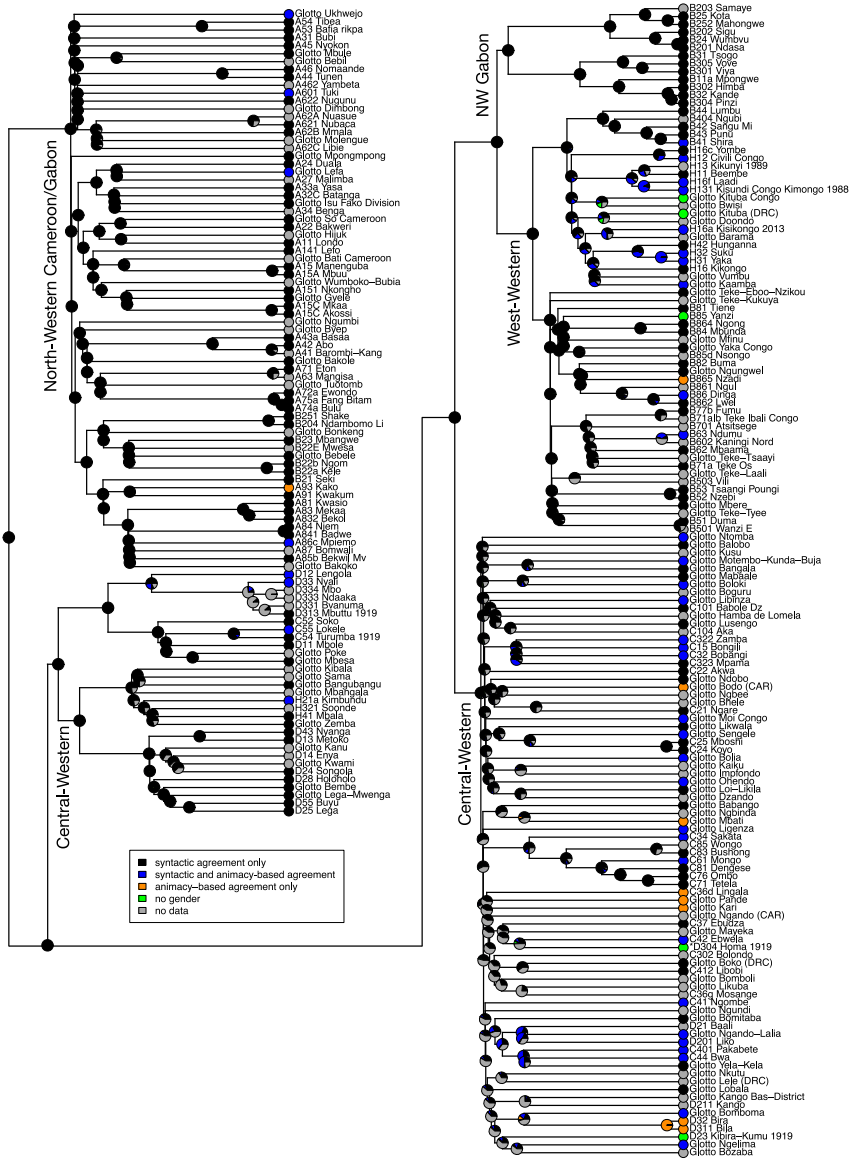

Figure 8: Gender systems of the languages of north-western Cameroon and Gabon. See Section 4.2 for the four-way typology displayed.

(Katamba 2003: 104ff). However, it is unclear at this point whether it also had animacy-based agreement. What we can observe though is that zone A languages, which typically are found in the highest nodes of the Bantu family tree and would thus have the greatest impact on the reconstruction of Proto-Bantu, mostly have only syntactic agreement in our sample. Thus assuming, as we do here, that Proto-Bantu might have only had syntactic agreement is no more than a conservative stance based on observational data.[30]

Figure 8 shows that systems attested in the languages of the sample are not randomly distributed but follow some clear genealogical patterns.[31] The top left of Figure 8 presents the gender systems of the NWB languages of Cameroon and Gabon. These are languages from Guthrie zone A and B20 and mostly have only syntactic agreement. The bottom left shows part of the Central-Western languages, which are again mostly languages with only syntactic agreement. In the top right, we have more North-Western Gabon languages, all of which have syntactic agreement. Then follow two subgroups with more variation; West-Western (mid-right) and the rest of Central-Western (bottom right). In the West-Western group, we find a subgroup containing most of the Guthrie zone H languages. Most of these languages have both syntactic and animacy-based agreement and this configuration can also be reconstructed for their most recent common ancestors. Most of the languages with only animacy-based gender or no gender are included in the Central-Western subgroup (bottom right). What is most important is that languages with restructured gender systems are clustered in small groups of related languages, which suggests that closely related languages are likely to have the same type of gender system.

The map (Figure 7) and the tree-based reconstructions (Figure 8) both show dependencies that require further explanation. Given the frequency of animacy-based agreement across NWB (51 out of 179 languages in our sample have some form of animacy-based agreement), as well as in other Bantu groups (Wald 1975), it seems that developing some form of animacy-based agreement comes naturally to

30 This assumption matches classical Proto-Bantu historical reconstructions going back to Meeussen (1967). See also Bostoen (2019: 322) and Van de Velde (2019) for more recent discussions.
31 This discussion is supported by applying Fritz and Purvis' phylogenetic signal test for binary data (Fritz and Purvis 2010). The data were binarized such that languages with only syntactic agreement formed one group, while the other group captured the three remaining types, i.e., languages with both syntactic and animacy-based agreement, languages with only animacy-based agreement, and languages without gender. The same tree discussed above (Koile et al. Under review) was used for the analyses, which were run with the function *phylo.d* from R package *caper* (Orme et al. 2013) in R (R Core Team 2018). The phylogenetic signal estimate *D* is 0.31 and thus significantly different from 1. This implies that the two states of the variable (syntactic agreement only vs. any other type of systems) are not randomly distributed across the tree.

(NW) Bantu languages. We speculate here that animacy-based agreement may have emerged independently across different NWB subgroupings, through inheritance or borrowing, and that, once emerged, it may exist alongside syntactic agreement for long periods of time. However, in our sample, languages without gender or with solely animacy-based gender are only found in areas that are geographically close to or even border with non-Bantu languages. As mentioned above, this may point to language contact as a catalyst of radical gender restructuring and erosion, an idea that we develop further in Verkerk and Di Garbo (2022). In the next section, we provide an exploratory account of the diachronic scenario that may explain changes in the distribution of animacy-based agreement.

5 Towards a diachronic typology of restructuring in NWB gender systems

Considering how common it is for languages with syntactic agreement to also have (some form of) semantic agreement, it cannot be excluded that the distributional patterns we find attested in NWB are the result of independent parallel developments. However, we also notice that clusters of closely related languages within the sample may sometimes reflect the entire or a substantial portion of this spectrum of typological variation, from solely syntactic agreement to solely animacy-based gender or no gender. This suggests that the different types of attested systems may be diachronically related to each other. One such cluster occurs within zone A90 where the three closely related languages, Kwakum (A91, kwak1266), Kako, and Polri, display some noticeable differences in the typological make-up of their gender agreement system (see Figure 8).[32] Kwakum, as many languages in the area, only has syntactic agreement even though in a somewhat reduced form, with eight singular/plural pairings and only noun-phrase internal agreement on a handful of targets (some of the numerals, the genitive constructions but only for some nouns, and the possessive pronouns, see Belliard 2007). In Kako, gender is completely animacy-based, as shown in (12), while Polri is completely devoid of gender, as shown in (14). Wega (2012) attributes this tendency towards reduction, which in varying degrees can be observed in all three languages, to the influence of Gbaya, a neighboring Ubangi language characterized by animacy-based gender agreement.

32 Unfortunately, Polri is missing from Figure 8 as neither Grollemund et al. (2015) nor Koile et al. (Under review) include it in their analyses.

In line with these observations, the hypothesis that we put forward here is that, in the NWB context, solely syntactic agreement and no gender may represent the two extremes of a diachronic continuum of restructuring, with various configurations of animacy-based agreement in between the two. The scenario that we suggest, and which we discuss in detail in this section, would be as follows. First, a language may only have syntactic agreement, while additional animacy-based marking is introduced once/if one or more targets allow(s) for semantic agreement. Our data suggest that optional and/or non-pervasive animacy-based agreement most typically occurs only with animate nouns, which, independently of their lexical gender, all receive class 1/2 agreement, while inanimate nouns typically retain their lexical gender and correspondent patterns of syntactic agreement. This is also the most typical, and better known, pattern of animacy-based agreement in Bantu languages beyond the northwestern area (see Wald 1975). However, as has been shown for Lika, Mpiemo and Bibaka Ukhwejo in Section 4.1, animacy-based agreement may also extend to the domain of inanimate nouns, with one marker starting indexing agreement with inanimate nouns on all or some of the agreement targets. When this happens, animacy-based agreement may completely take over. It affects animate and inanimate nouns alike and, if it extends to all available agreement targets, no trace of syntactic agreement remains and gender marking becomes entirely animacy-based. We call this phenomenon *generalized animacy-based agreement*. If even animacy-based agreement is lost, no productive gender system remains, even though some fossilized remnants of gender marking may still survive on nouns. The suggested diachronic pathway can be summarised as follows:

(16)
1. only syntactic gender agreement >
2. syntactic and animacy-based agreement with animate nouns >
3. syntactic and animacy-based agreement with animate and inanimate nouns >
4. only animacy-based agreement >
5. no gender

While this proposal reflects earlier suggestions by Maho (1999: 127–142), we discuss the added explanatory power and empirical validity of our analysis in the remaining of this section. First of all, by suggesting that there may be a diachronic order to the restructuring of NWB gender systems, and that one of the triggers of restructuring is the spreading of animacy-based agreement, we do not intend to imply that all Bantu languages with optional animacy-based agreement are on a path towards loss of gender. On the contrary, as we also show in this article, there are many languages in our sample, and in other parts of the Bantu-speaking world (e.g., eastern coastal Bantu) where animacy-based agreement is restricted to

animate nouns and optionally manifested only on some of the agreement targets, while traditional patterns of gender marking are used with the majority of nouns and in the majority of syntactic contexts. In such cases, the coexistence of syntactic and animacy-based agreement can be described as a stable pattern of variation which can remain unchanged for centuries.

However, we find that animacy-based agreement has the potential of becoming a major trigger of restructuring in NWB gender systems when (1) it spreads to an increasingly high number of agreement targets, (2) it becomes obligatory (at least for some nouns and/or in certain syntactic contexts), and (3) it extends to inanimate nouns.

Observation (1) and (2) are not new. For instance, Section 8.3 of Corbett (1991) is entirely devoted to discuss how the spreading of semantic agreement along the lines of the Agreement Hierarchy may lead to substantial changes to both gender assignment and gender agreement. By using a wealth of examples from a variety of Bantu, other Atlantic-Congo, and European languages, Corbett shows how these changes essentially hinge upon two major diachronic processes, one whereby semantic agreement may be gradually generalized to all available agreement targets (starting with pronouns and finishing off with attributive modifiers), and one whereby an increasingly high number of nouns obligatorily select semantic agreement, ultimately causing a reshuffling in gender assignement rules: "if small numbers of nouns are involved the effect on the system will be negligible, but if several nouns follow the same path, then the assignment system itself may change" (Corbett 1991: 248). According to Corbett, a case in point to illustrate both processes are the north-east coastal Bantu languages studied by Wald (1975), where, as mentioned in Section 2.3, a whole range of variation in terms of degrees of pervasiveness (how many agreement targets) and obligatoriness of animacy-based agreement is attested. In Wald's sample, Bondei represents the end point of this typological continuum. In Bondei, animacy-based agreement is obligatory with animate nouns and on all agreement targets, which means that all animate nouns are assigned to gender 1/2, while inanimate nouns continue being assigned to the many different genders that the language retains (for an overview of gender distinctions in Bondei see Merlevede 1995).

The third observation, *generalized animacy-based agreement*, has, to the best of our knowledge, never been brought to the fore before. It entails that the spreading of animacy-based agreement may lead to a reduction in the number of gender distinctions when it extends to the domain of inanimate nouns. This development could be explained in terms of known generalizations about animacy effects in the spreading of language change. It is a well-established fact in general linguistic and typological literature that the spreading of patterns of variation and change having scope on nominal morphosyntax may be lexically constrained

along the lines of the Animacy Hierarchy (Corbett 2000; Dahl and Fraurud 1996; Enger and Nesset 2011). Variation and change may start off with animate nouns and later expand to the inanimates, what Enger and Nesset (2011) refer to as a *top-down* type change, but the inverse direction, from inanimate to animate nouns, what Enger and Nesset (2011) refer to as a *bottom-up* change, is also possible. The patterns of variation and change that we observe in a minority of languages of the sample with respect to the spreading of animacy-based agreement would be of the top-down kind, in that animacy-based agreement first affects only animate nouns and later spreads to the inanimates. According to Enger and Nesset (2011), this type of path is fairly typical for animacy-driven diachronic change in the domain of gender marking. While the diachronic evidence needed in order to fully confirm the validity of this proposal is currently not available given the status of description of many of the relevant languages of the area, we believe that Lika, Mpiemo and Bibaka Ukhwejo (discussed in Section 4.1) offer some evidence in support of this suggestion.

In these three languages, gender distinctions in the domain of inanimate nouns have become or are in the process of becoming neutralized in that, similarly to animate nouns, inanimate nouns become associated with only one agreement class. Syntactic agreement coexists with instances of animacy-based agreement and both animate and inanimate nouns undergo animacy-based agreement. As mentioned in Section 4.1, subject-verb agreement in Lika only distinguishes between animate and inanimate nouns. Animate subjects take prefix *a-/ø* in the singular and prefix *ba*-in the plural, while inanimate subjects take *a/o* both in the singular and plural (cf. Example 11-d). Nouns still retain their lexical gender, which is marked on adnominal modifiers. Similarly to Lika, both in Mpiemo and Bibaka Ukhwejo restructuring and reduction in number of gender distinctions appear to be connected to the generalization of animacy-based agreement to both animate and inanimate nouns through the use of agreement pattern 1/2 for the former, and 7/8 (Mpiemo) or just 7 (Bibaka Ukhwejo) for the latter type of nouns. What differentiates Lika from Mpiemo and Bibaka Ukhwejo is the fact that in the latter two languages, generalized animacy-based agreement is reported for all agreement targets, while in Lika it only affects subject agreement on verbs.

We also find evidence for generalized animacy-based agreement in the phonological shape of agremeent markers. Two closely related languages Bwa and Pagibete (both spoken in Congo and the Democratic Republic of the Congo) display neutralization of gender distinctions in the domain of subject agreement. They both retain instances of syntactic agreement on other targets, but do not have any form of gender agreement on verbs, where they only differentiate between singular and plural subjects. In both languages, singular subjects take subject agreement

prefix *a*-while the plurals take subject prefix *ba*- (Motingea Mangulu 2005; Reeder 1998).[33] The shape of these markers is clearly reminiscent of class 1 and 2 agreement markers. This suggests that the neutralization of gender distinctions in the domain of subject agreement in these languages might have come about through the overextension of the agreement pattern associated with animate nouns to the inanimates, a development which is similar, albeit in the opposite direction, to what is currently ongoing in Bibaka Ukhwejo, where the inanimate agreement prefix is generalized to all contexts. Thus, while loss of gender distinctions in these languages is restricted to only one agreement target, it serves as an illustration of the morphosyntactic processes which, if extended to other targets, may lead to further erosion of gender marking.[34]

Generally speaking, in languages where gender distinctions are either partially or completely neutralized, or which exhibit solely animacy-based gender systems, the markers that are used to encode animacy and/or number distinctions are often reminiscent of the markers that are typically recruited for the purpose of animacy-based agreement in languages with more conservative systems. This could suggest that, before undergoing further restructuring and/or loss, these languages may have also gone through a stage of optional animacy-based agreement. In languages with solely animacy-based gender, for instance, markers that are clearly reminiscent of classes 1 and 2 are typically used with animate nouns, as in Bera where the prefixes *mu*- and *ba*-are used as nominal and agreement markers of singular and plural animate nouns, respectively (Susa 1972). The morphological realization of nominal and agreement marking with inanimate nouns tends to be more varied in the languages of our sample with solely animacy-based gender. Thus, no comprehensive account can be given based on the data at hand, which are often scanty and hard to grasp from a comparative perspective. We also note that in languages that have completely lost gender, former class 2 prefix *ba*-may still be used as a general nominal pluralizer, as in Komo where no gender agreement is left, but *ba*-can be used to pluralize any animate noun (Thomas 1994).

33 For Bwa, Motingea Mangulu (2005: 36) also mentions that even though the singular/plural distinction is the only productive grammatical distinction that is marked through subject agreement, animate/inanimate marking (with *i*-used for agreement with inanimate subjects) may sometimes be used. Since, however, this is described as a less frequent pattern of encoding and no clear illustrations are given, we coded Bwa as lacking animate agreement in the domain of subject agreement.

34 Another case in point would be Bongili (C15, bong1284). Bongili still retains animate agreement on verbs (and syntactic agreement on a variety of targets), but may also mark subject agreement on the verb through the number- and gender-invariant prefix *a*-, which is historically the verbal agreement marker of class 1 (Motingea Mangulu 2008).

The last step of the diachronic pathway proposed in 5 would suggest that when and if generalized animacy distinctions are lost, no productive gender system survives. While no language in our dataset fully illustrates this pattern of development (from animacy-based gender to no gender), data from a handful of the sampled languages give us a sense of how a process of this type may unfold. As mentioned above, Thornell (2012) notices that some Bibaka Ukhwejo speakers tend to overextend the use of agreement prefix *y-*, typically associated with inanimate nouns, to all nouns and in all syntactic contexts in which gender agreement would surface. While there is considerable inter- and intraspeaker variation in the usage of generalized *y-*, the spreading of this pattern, which is in turn diachronically connected with generalized animacy-based agreement, may eventually lead to the complete loss of gender in the language (Thornell 2012).

We do not suggest that the diachronic pathway described and illustrated in this section is the sole process leading to the restructuring of NWB gender systems, let alone of the gender systems of the larger Bantu family. We can easily imagine alternatives, for instance, the complete loss of syntactic agreement may be followed by the later re-emergence of an animacy-based gender system or by animacy-based number marking. We also have Shiwa in our sample, where restructuring and loss of gender distinctions is not related to the spreading of any form of semantic agreement (see Section 4.1). Generally speaking, slow-moving diachronic change and ancestry are clearly not the only factors at stake in explaining gender restructuring in many of the languages of the sample, such as the creole languages Kituba and Kinshasa Lingala, or the northern borderland languages Mbati and Pande. As mentioned in Section 4.4, other driving forces of change, related to language contact and population history, should be factored in. While the proposed diachronic pathway would presuppose a chain of gradual changes that may fit a sociolinguistic scenario of prolonged bilingualism and long-term language contact between diverse speech communities, we cannot exclude that more abrupt changes, related to rapid language shift or pidginization, play an equally important role.[35] However, lack of diachronic data, both on gender systems and on sociological characteristics of speech communities, makes it hard to find evidence for the proposed pathway beyond what has been mentioned above.

An additional dimension of analysis is the relationship between gender restructuring and the number of targets exhibiting syntactic versus animacy-based

35 Both types of processes, gradual and abrupt change, and in relationship to both types of contact scenarios, long-term contact versus abrupt language shift, have been reported in recent literature on contact-induced change in the domain of gender marking. See Di Garbo (2020) for an overview.

agreement. As mentioned before, in the languages of our sample, there is a tendency for syntactic agreement to occur on no less than five agreement targets. Moreover, in languages where both agreement patterns are attested, animacy-based agreement tends to be less pervasive than syntactic agreement. However, both Bibaka Ukhwejo and Mpiemo, where gender marking is undergoing erosion, have less than five targets agreeing syntactically, and all of them can in principle also carry animacy-based agreement. We find 10 additional languages that only mark syntactic agreement on five or fewer targets. Three of these 10 languages are spoken in close proximity to languages with highly reduced gender systems. These are Songoora (D24, song1300), a close neighbor of Komo which does not have gender; Bekwil (A85b, bekw1242), closely related to Mpiemo, which is undergoing heavy restructuring and possibly also erosion of gender marking; and Kwakum, which is closely related to Polri, also a genderless language. While, similarly to their neighbors, these three languages, Songoora, Bekwil and Kwakum, may also be on their way to lose and/or restructure their gender system, the sources at hand do not give any hints that this is indeed the case. Mbangwe (B23, mban1268), Ndasa (sud) (B201, ndas1238), Ngom (nord) (B22b, ngom1270), and Wumbvu (B24, wumb1242) have syntactic agreement on five targets and are all closely related to each other, but, as far as we can tell, they are not neighbors with languages with restructured or eroded systems. For the remaining three languages, Mbule (A623, mbul1262), Ombamba (B62, omba1241), and Nyokon (A45, nyok1243), there is no immediate reason why they should have only three or four targets of syntactic agreement. These latter facts may indicate that less pervasive systems of gender marking, where syntactic agreement is only marked on a small number of targets, are also a part of the typological spectrum of variation in NWB, independently of animacy-based agreement and without necessarily signaling ongoing erosion. Nevertheless, we find that the relationship between number of agreement targets and gender restructuring and/or loss would deserve to be further investigated.

To conclude, while we find the proposed diachronic pathway from solely syntactic agreement to no gender suggestive, and we think that the observations gathered in this section support it for at least some of the languages of the sample, we leave its affirmation to future research.

6 Concluding remarks

In this article, we pulled together two phenomena previously discussed in the typology of Bantu gender systems, animacy-based agreement and highly reduced gender systems, and showed how these may be related on a continuum of increasing influence of animacy-based restructuring. Animacy-based agreement is

certainly widespread in NWB, where at least 40 languages out of 179 have both syntactic and animacy-based agreement and 11 languages have only animacy-based gender. The picture emerging from the NWB data thus calls into question the generalizations of earlier studies (Contini-Morava 2008), where animacy-based agreement was described as a peculiar feature of eastern Bantu languages. In addition, given that animacy-based agreement is considered to be under-reported in grammars (Maho 1999), our findings based on reference grammars are rather surprising and call for further hypothesis testing in other areas of the Bantu-speaking world, both at the descriptive and comparative level. While Bantu languages are usually portrayed as a solid block of conservative gender systems, NWB languages provide us with a more fine-grained picture of the range of variation that is found in the family in this domain of grammar. Whether any of the patterns uncovered in this study stretches beyond this area, i.e., towards the eastern and southern Bantu languages, remains to be seen. In this sense, we find the prevalence of animacy-based agreement in the languages of zone H, the southernmost languages investigated in this study, highly suggestive. Wald (1975) of course also finds various types of animacy-based agreement in eastern Bantu.

In Section 4.2, Figure 3, we propose a four-way categorical typology of NWB gender systems by cross-tabulating the targets that receive syntactic agreement and those that inflect for animacy. This typology distinguishes between languages with solely syntactic agreement, languages with syntactic and animacy-based agreement, languages with solely animacy-based gender, and languages with no gender. According to our data, these four types have characteristic distributions of the number of targets that inflect for either or both types of agreement. However, as the qualitative overview in Section 4.1 also shows, it is important to stress that the differences between languages from different types can be really small, and are best conceived of as a continuum. In addition, alternative parameters of classification, such as for instance the obligatoriness of animacy-based agreement, should be applied when the data allows.

The proposed typology also matches earlier observations by Maho (1999) on crosslinguistic variation in Bantu gender systems, which we summarized in Section 2.3, Table 1. The languages of our sample largely align with the types suggested by Maho. Interestingly, they also confirm some of the gaps he found in the range of attested logically possible types. For instance, in our data set we do not find any language which would only mark number distinctions both on nouns and through agreement (Type D4 in Maho's typology). Our sample includes languages like Pande, where nouns only inflect according to number but trigger animacy-based agreement, and Polri, where the opposite system is attested, with nouns carrying some relics of animacy-based inflections, but only triggering number agreement. Languages that would only mark number distinctions through both

nominal inflections and agreement patterns do not occur in this data set. While this could of course be a matter of chance, it also suggests that animacy distinctions may be an highly entrenched feature of the languages of the area, Bantu and non-Bantu alike, possibly as a result of substrate influence from autochtonous languages (see Section 4.4). This observation is no more than speculative at this stage, and could only be tested via a comprehensive study of agreement systems across the entire Bantu-speaking world, supported by systematic comparisons with the agreement systems attested in the respective contact languages. It should also be stressed that teasing apart areal patterns from genealogical innovations and retentions in the Bantu family is generally a hard task, complicated by the fact that close neighbors are also often closely related sister languages.

Coming to the rise and spread of animacy-based agreement, our findings align with existing typological generalizations, which predict that semantic agreement encroaches syntactic agreement along the lines of the Agreement Hierarchy (Corbett 1979, 1991, 2000). We found that the most frequent hosts of animacy-based agreement in NWB languages are the markers of subject agreement on the verb and the third person pronouns, which further reinforces the idea that anaphoras are the most likely attractors of semantic agreement crosslinguistically. We also found evidence in support of the existence of a hierarchy of syntactic integration between nouns and different adnominal modifiers, which manifests itself through the fact that demonstratives, possessives and numerals are more resistant to animacy-based agreement than other types of adnominal modifiers such as adjectives. These findings are new, and open up the possibility of expanding and further detailing the predictions entailed by the Agreement Hierarchy. They also suggest that looking at the inflections carried by individual agreement targets, as we do here, rather than focusing on agreement domains as a whole, as posited in the hierarchy, is a very promising way of uncovering more fine-grained hierarchical effects related to linear distance between controller nouns and agreement hosts. More studies, within Bantu and beyond, should be conducted in order to validate these suggestions further.

Finally, in line with previous generalizations on animacy effects in the diachrony of gender systems, we found that in most cases, animacy-based agreement only affects the top left end of the Animacy Hierarchy, that is animate nouns. However, in some languages, animacy-based agreement also spreads to the domain of inanimate nouns, what we call generalized animacy-based agreement. In Section 5, we suggest that generalized animacy-based agreement could be one of the mechanisms that paves the way to the highly eroded gender systems attested in the languages of the sample, where animacy distinctions are the only type of

distinction encoded through gender assignment and agreement. These findings are also new and enrich state-of-the-art knowledge on the typology and evolution of Bantu nominal morphosyntax. Yet, it should be stressed that only a handful of the sampled languages currently bring support to the suggested diachronic pathway (from syntactic agreement to animacy-based gender and, eventually no gender). Further studies, spanning the rest of the Bantu family, are needed both at the descriptive and comparative level.

Acknowledgments: This article is the result of a research collaboration begun in 2016, during the first Quantitative Methods Spring School, organized by the Department of Linguistic and Cultural Evolution at the Max Planck Institute for the Science of Human History (Jena). We thank the spring school's directors, Russell Gray and Fiona Jordan, and all participants for support and inspiration. Ongoing analyses and preliminary results of this research have been presented at various conferences, workshops, and other academic venues. We are particularly thankful to the audiences of the workshop on "Language shift and substratum interference in (pre)history" (Jena 2017), the SLE conferences of 2017 (Zürich) and 2018 (Tallinn), and the African Linguistics research seminar of the Humboldt University (Berlin, 2017). We wish to thank two anonymous reviewers for constructive comments and criticism, Volker Gast and Ann Kelly for editorial assistance, Ines Fiedler, Tom Güldemann, Brigitte Pakendorf, Kaius Sinnemäki, Bernhard Wälchli, and Ricardo Napoleão de Souza for feedback and stimulating discussions. For help with data collection, we are thankful to Harald Hammarström and the Royal Museum of Central Africa. The usual disclaimers apply.
Research funding: For this research Di Garbo has received funding from the Wenner-Gren Foundations (Sweden) and partly from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 805371).
Data availability: The data generated and analyzed during this study are available in the Zenodo repository: https://doi.org/10.5281/zenodo.6378548.

Appendix A: Coding model

The following coding model has been used for all languages included in the study.

The two first sets of questions aim to gather information about the inventory size of noun class forms and agreement classes.

1. Gender marking on nouns
   - how many singular noun class forms?
   - how many plural noun class forms?
   - how many number-invariant noun class forms?
   - how many singular/plural pairings of noun class forms?
2. Gender marking on agreement targets
   - how many distinguishable singular agreement classes?
   - how many distinguishable plural agreement classes?
   - how many number-invariant agreement classes?
   - how many paired singular/plural agreement classes?

Question 3 and 4 are concerned with the distribution of syntactic and animacy-based agreement and read as follows:

3. What are the word classes that carry syntactic agreement?
4. What are the word classes that carry animacy-based agreement?

In order to answer these questions we investigated presence and/or absence of one or the other type of agreement pattern based on a set of 15 targets (14 different word classes plus a category "other"). In the following, we provide a list of target types as well the definitions we used in order to identify them across languages. The same inventory of target types is used to answer both question 3 and 4.

A.1: List of target types used to answer question 3 and 4

For each of the two questions, and with respect to each and every target type, variable coding is "Yes/No/No data" (e.g.: "Do attributive modifiers agree syntactically? Yes/No/No data", and "Do attributive modifiers agree semantically, i.e., based on animacy? Yes/No/No data"). Except for the variable "other", which is listed at the end, variable names are ordered alphabetically.

Definitions of each target type are complemented with illustrations taken from the languages of our sample. These examples are meant to illustrate hosts of gender marking in different word classes and/or types of constructions. Even though language-specific hosts and patterns of gender marking largely differ across languages, the examples we provide can be easily generalized across languages since gender marking always involves prefixation on a given agreement target. For convenience sake, most of the examples come from Dibole, a language with only syntactic agreement, as described by Leitch (2003). Those target types which are not described for Dibole are exemplified through other languages.

- **Attributive adjectives**: adnominal modifiers encoding property words. Example from Dibole: *-bé* 'bad', *-lámú* 'good' (Leitch 2003: 418).

- **Copula-like constructions**: constructions expressing nominal and/or locative predications. Example from Dibole: *-é* (Kutsch Lojenga 2003: 418); the copula verb can also be omitted, in which case it is the noun or the adjective that carries the agreement marker (Leitch p.c.).
- **Demonstrative modifiers**: adnominal modifiers indicating different degrees of spatial distance from the speaker and/or the listener. Example from Dibole: *-ò* 'Proximate I'; *-wá* 'Distant I' (Leitch 2003: 416).
- **Demonstrative pronouns**: pronominal expressions indicating different degrees of spatial distance from the speakers and/or the listener. Often the same as adnominal modifiers. Example from Dibole: *- ò* 'Proximate I'; *-wá* 'Distant I' (Leitch 2003: 416–417).
- **Genitives/connectives**: in Bantu languages, these are typically markers that are used to introduce nominal possessors. They generally consist of the stem *a* preceded by a pronominal prefix, which agrees in gender with the possessor. They are also used to encode adjectival type of meanings with modifying nouns encoding properties and/or entities. Example from Dibole: *ò -à*, genitive marked by class 1 prefix (Leitch 2003: 419).
- **Independent third person pronouns**: anaphoric pronouns corresponding to 'he/she/it' in English. Example from Dibole: *-angò/-angoá* 'it/them' (Leitch 2003: 417).
- **Numerals**: adnominal modifiers encoding cardinal numbers. Ordinal numbers also agree in gender in Bantu languages, but they are expressed through genitive constructions with cardinal numbers as modifiers (thus gender agreement is marked on the genitive relator rather than on the numeral as such). Example of cardinal number from Dibole: *-hɔ́kɔ́* 'one' (Leitch 2003: 417).
- **Quantifiers**: adnominal modifiers encoding quantity expressions such as, for instance, 'some', 'all', 'many'. Example from Dibole: *-esú* 'all' (Leitch 2003: 417).
- **Possessive pronouns**: pronominal expressions agreeing in gender with the possessee, corresponding to the English 'my/your/his/her/our/their'. In Bembe, possessive pronouns are formed by attaching a gender agreement marker to a possessive root. The language distinguishes six possessive roots, one for each person and number value (first/second/third person and singular/plural number). Example: *-ane* 'my', *-obe* 'your' (Iorio 2011: 55).
- **Predicative adjectives**: property words used predicatively. Example from Dibole: *-bé* 'bad', *-lámú* 'good', with the copula being also inflected if present (Leitch 2003: 418).

- **Question words**: selective interrogative such as 'how many?' and 'which?', as well as interrogative pronouns ('who?' 'what?'). Examples from Dibole: *-sò* 'which thing?'; *ndzá* 'who?/what?' (Leitch 2003: 416). We are aware of the fact that non-selective interrogative pronouns do not qualify as agreement targets in the proper sense of the term, given that among other properties, their referential specification is, by definition, unknown/suspended (Idiatov 2007). Nevertheless, we chose to include them in our inventory of syntactic hosts because we were interested in capturing how often basic animacy-based contrasts are coded in this domain across the languages of the sample. Selective interrogative pronouns in Bantu are, on the other hand, gender agreement targets in the proper sense of the term and often inflect based on the lexical gender of the noun they substitute for.
- **Reflexive pronouns**: reflexives in Bantu are usually invariable prefixes, which are part of the set of inflectional markers that a verb can take. Example from Dibole: *-á-* (Leitch 2003: 416). In some cases we find reflexive intensifiers, which are independent words that can take pronominal markers in agreement with the gender of the noun. Example from Basaá: *mdɛ́* is added to the independent pronoun, which is in turn inflected for gender (Hayman 2003: 19–20).
- **Relative clauses**: in Bantu languages, relative clauses may be formed in a variety of ways: through relative markers, which are affixed to the verb and agree in gender with the head of the relative clause, through the use of associative markers that agree in gender with the head of the relative clause, or through the use of demonstratives also agreeing in gender with the head of the relative clause. We try to capture all of these patterns when looking at relative constructions in the languages of our sample. Example of the first type: in Dibole, if the subject of the relative clause is a full noun phrase, relativization is marked through tonal downstep on the verbal argument prefix which agrees with the head of the relative clause (Leitch 2003: 420). Example of the second type: in Dibole, if the subject of the relative clause is a pronoun, relativization is marked through an associative marker which agrees in gender with the head of the relative clause (Leitch 2003: 421). Example of the third type: in Tuki, gender-marked demonstratives (e.g., *-jó*, 'this one') can be used to introduce relative clauses (Hyman 1980: 34).
- **Verbs**: lexemes for the encoding of prototypical predicative expressions (actions, states). In our coding design, presence of syntactic and/or animacy-based agreement on the verb means either marking of the subject or the object. Example from Dibole: *-dzé* 'eat' (Leitch 2003: 420).
- **Other targets and/or domains of gender marking**: Here we include anything that cannot be captured by the features listed above. Example from Budza: *-múíni* 'themselves', not a reflexive marker but rather a contrastive modifier, i.e., 'the chiefs themselves went up and …' (Stappers 1955: 108).

A.2: Additional questions

- Is animacy-based agreement obligatory outside the NP?
- Is animacy-based agreement obligatory everywhere?
- Does agreement only signal number?
- Do noun class forms only mark number?
- Do noun class forms only mark animacy?
- Do noun class forms mark animacy and number?
- Is there extra-marking of animacy on nouns (e.g., animacy markers are juxtaposed to the traditional nominal gender markers)?
- Is there extra-marking of plurality on nouns (e.g., in addition to their traditional gender/number markers, nouns take an additional plural marker which is gender-invariant)?
- Is there extra-marking of animacy and number on nouns (e.g., animacy/number markers are juxtaposed to traditional gender/number markers?)
- Notes (this is a free text variable where the coder can add any additional remark on the language which is being described).

Appendix B: The languages of the sample

Name	Isocode	Glottocode	Guthrie	Data coverage
Akoose	bss	akoo1248	A15	Yes
Akwa	akw	akwa1248	C22	Yes
Amba (Uganda)	rwm	amba1263	D22	Yes
Babango	bbm	baba1263	C441	Yes
Bafaw-Balong	bwt	bafa1247	A141	Yes
Bafia	ksf	bafi1243	A53	Yes
Bafoto	–	bafo1235	C611	Yes
Bakaka	bqz	baka1273	A15	Yes
Bakoko	bkh	bako1249	A43b	No
Bakole	kme	bako1250	A231	Yes
Bali (DRC)	bcp	bali1274	D21	No
Baloi	biz	balo1261	C31	Yes
Bamwe	bmg	bamw1238	C412	Yes
Bangala	bxg	bang1353	C30A	Yes
Bangi	bni	bang1354	C32	Yes
Bangubangu	bnx	bang1350	D27	Yes
Bankon	abb	bank1256	A42	Yes
Barama	bbg	bara1362	B402	No
Barombi	bbi	baro1252	A41	No

(continued)

Name	Isocode	Glottocode	Guthrie	Data coverage
Basa (Cameroon)	bas	basa1284	A43a	Yes
Bassossi	bsi	bass1260	A15	Yes
Batanga	bnm	bata1285	A32	Yes
Bati (Cameroon)	btc	bati1251	A65	No
Bebele	beb	bebe1248	A73a	Yes
Bebil	bxp	bebi1242	A73b	No
Beeke	bkf	beek1238	D335	Yes
Beembe	beq	beem1239	H11	Yes
Bekwil	bkw	bekw1242	A85b	Yes
Bembe	bmb	bemb1255	D54	Yes
Benga	bng	beng1282	A34	No
Bera	brf	bera1259	D32	Yes
Bhele	bhy	bhel1238	D31	No
Bila	bip	bila1255	D311	Yes
Bodo (CAR)	boy	bodo1272	D308	Yes
Boguru	bqu	bogu1241	D302	No
Boko (DRC)	bkp	boko1263	C16	Yes
Bolia	bli	boli1255	C35	Yes
Bolo	blv	bolo1261	H23	No
Boloki	bkt	bolo1262	C36e	Yes
Bolondo	bzm	bolo1263	C302	No
Boma	boh	boma1246	B82	Yes
Bomboli	bml	bomb1261	C161	No
Bomboma	bws	bomb1262	C411	Yes
Bomitaba	zmx	bomi1238	C14	Yes
Bomwali	bmw	bomw1238	A87	No
Bongili	bui	bong1284	C15	Yes
Bonjo	bok	bonj1234	C10?	No
Bonkeng	bvg	bonk1243	A14	No
Bozaba	bzo	boza1238	C162	No
Bube	bvb	bube1242	A31	Yes
Bubi	buw	bubi1250	B305	Yes
Budu	buu	budu1250	D32	Yes
Budza	bja	budz1238	C37	Yes
Bulu (Cameroon)	bum	bulu1251	A74	Yes
Bushoong	buf	bush1247	C83	Yes
Buyu	byi	buyu1239	D55	Yes
Bwa	bww	bwaa1238	C44	Yes
Bwela	bwl	bwel1238	C42	Yes
Bwisi	bwz	bwis1242	B401	No
Byep	mkk	byep1241	A831	No
Dengese	dez	deng1250	C81	Yes
Dibole	bvx	dibo1245	C101	Yes
Dimbong	dii	dimb1238	A52	No

(continued)

Name	Isocode	Glottocode	Guthrie	Data coverage
Ding	diz	ding1239	B86	Yes
Doondo	dde	doon1238	H112B	No
Duala	dua	dual1243	A24	Yes
Duma	dma	duma1253	B51	Yes
Dzando	dzn	dzan1238	C413	No
Elip	ekm	elip1238	A62	No
Enya	gey	enya1247	D14	No
Eton (Cameroon)	eto	eton1253	A71	Yes
Ewondo	ewo	ewon1239	A72	Yes
Fang (EG)	fan	fang1246	A75	Yes
Gyele	gyi	gyel1242	A801	Yes
Hamba (DRC)	hba	hamb1245	C71	No
Hijuk	hij	hiju1238	A501	No
Holoholo	hoo	holo1240	D28	Yes
Homa	hom	homa1239	D304	Yes
Hungana	hum	hung1278	H42	Yes
Ibali Teke	tek	ibal1241	B75	No
Isu (Fako Division)	szv	isuf1235	A23	Yes
Kaamba	xku	kaam1238	H112A	Yes
Kaiku	kkq	kaik1247	D312	No
Kako	kkj	kako1242	A93	Yes
Kande	kbs	kand1300	B32	Yes
Kango (Bas-Uélé District)	kty	kang1286	C403	No
Kango (Tshopo District)	kzy	kang1285	D211	No
Kaningi	kzo	kani1279	B602	No
Kanu	khx	kanu1278	D251	No
Kari (DRC)	kbj	kari1306	D301	Yes
Kélé (Gabon)	keb	kele1257	B22	Yes
Kele-Foma (DRC)	khy	kele1255	C55	Yes
Kimbundu	kmb	kimb1241	H21	Yes
Kituba (Congo)	mkw	kitu1245	H10B	Yes
Kituba (DRC)	ktu	kitu1246	H10A	Yes
Kol (Cameroon)	biw	kolc1235	A832	Yes
Komo (DRC)	kmw	komo1260	D23	Yes
Koongo	kng	koon1244	H14	Yes
Koonzime	ozm	koon1245	A842	Yes
Kota (Gabon)	koq	kota1274	B25	Yes
Koyo	koh	koyo1242	C24	Yes
Kunyi	njx	kuny1238	H13	No
Kusu	ksv	kusu1252	C72	No
Kwakum	kwu	kwak1266	A91	Yes
Kwami	ktf	kwam1250	D43	No
Kwasio	nmg	kwas1243	A81	Yes
Laari	ldi	laar1238	H16f	Yes

(continued)

Name	Isocode	Glottocode	Guthrie	Data coverage
Lefa	lfa	lefa1242	A51	Yes
Lega-Mwenga	lgm	lega1250	D25	Yes
Lega-Shabunda	lea	lega1249	D251	Yes
Lele (DRC)	lel	lele1265	C84	No
Lengola	lej	leng1258	D12	Yes
Leti (Cameroon)	leo	leti1245	A601	No
Libinza	liz	libi1244	C321	Yes
Ligenza	lgz	lige1238	C414	Yes
Lika	lik	lika1243	D201	Yes
Likila	lie	liki1240	C31	Yes
Likuba	kxx	liku1242	C27	No
Likwala	kwc	likw1239	C26	Yes
Lingala (Kinshasa)	lin	ling1263	C30b	Yes
Lobala	loq	loba1239	C16	Yes
Lombo	loo	lomb1260	C54	Yes
Lumbu	lup	lumb1249	B44	Yes
Lusengo	lse	luse1252	C36	Yes
Lwel	–	lwel1234	B85	Yes
Mabaale	mmz	maba1270	C31	Yes
Mahongwe	mhb	maho1248	B252	Yes
Makaa	mcp	maka1304	A83	Yes
Malimba	mzd	mali1280	A27	No
Mayeka	myc	maye1238	D307	No
Mbala	mdp	mbal1257	H41	Yes
Mbangala	mxg	mban1264	H34	No
Mbangwe	zmn	mban1268	B23	Yes
Mbati	mdn	mbat1248	C13	Yes
Mbere	mdt	mber1257	B61	Yes
Mbesa	zms	mbes1238	C51	Yes
Mbo (Cameroon)	mbo	mboc1235	A15	Yes
Mbo (DRC)	zmw	mbod1238	D334	No
Mboko	mdu	mbok1243	C21	Yes
Mbole	mdq	mbol1247	D11	Yes
Mbosi	mdw	mbos1242	C25	Yes
Mbule	mlb	mbul1262	A623	Yes
Mfinu	zmf	mfin1238	B83	No
Mituku	zmq	mitu1240	D13	Yes
Mmaala	mmu	mmaa1238	A62	Yes
Moi (Congo)	mow	moic1236	C32	Yes
Mokpwe	bri	mokp1239	A21	Yes
Molengue	bxc	mole1238	B221	No
Mongo	lol	mong1338	C61	Yes
Mpama	–	mpam1239	C323	Yes
Mpiemo	mcx	mpie1238	A86c	Yes

(continued)

Name	Isocode	Glottocode	Guthrie	Data coverage
Mpongmpong	mgg	mpon1254	A86	Yes
Mpuono	zmp	mpuo1241	B84	Yes
Mwesa	–	mwes1234	B22E	No
Myene	mye	myen1241	B11	Yes
Ndaka	ndk	ndak1241	D333	No
Ndambomo	nxo	ndam1254	B204	Yes
Ndasa sud	nda	ndas1238	B201	Yes
Ndobo	ndw	ndob1238	C31	Yes
Ndolo	ndl	ndol1238	C36g	No
Ndumu	nmd	ndum1239	B63	Yes
Ngando (CAR)	ngd	ngan1304	C15	No
Ngando (DRC)	nxd	ngan1302	C63	Yes
Ngbee	jgb	ngbe1238	D336	No
Ngbinda	nbd	ngbi1238	D303	No
Ngelima	agh	ngel1238	C45	Yes
Ngom nord	nra	ngom1270	B22	Yes
Ngombe (DRC)	ngc	ngom1268	C41	Yes
Ngongo (DRC)	noq	ngon1267	H31	Yes
Ngubi	–	ngub1239	B404	No
Ngul	nlo	ngul1247	B86	No
Ngumbi	nui	ngum1255	A33b	No
Ngundi	ndn	ngun1270	C11	No
Ngungwel	ngz	ngun1272	B72a	Yes
Njebi	nzb	njeb1242	B52	Yes
Njyem	njy	njye1238	A84	Yes
Nkongho	nkc	nkon1247	A151	Yes
Nkutu	nkw	nkut1238	C73	No
Nomaande	lem	noma1260	A46	Yes
Nsongo	nsx	nson1238	H24	No
Ntomba	nto	ntom1248	C35	Yes
Nubaca	baf	nuba1241	A621	Yes
Nugunu (Cameroon)	yas	nugu1242	A622	Yes
Nyali	nlj	nyal1250	D33	Yes
Nyanga	nyj	nyan1304	D43	Yes
Nyanga-li	nyc	nyan1303	D305	No
Nyokon	nvo	nyok1243	A45	Yes
Nzadi	–	nzad1234	B85	Yes
Ombamba	mbm	omba1241	B62	Yes
Ombo	oml	ombo1238	C76	Yes
Oroko	bdu	orok1266	A101	Yes
Osamayi	syx	osam1235	B203	No
Pagibete	pae	pagi1243	C401	Yes
Pande	bkj	pand1264	C12	Yes
Pinji	pic	pinj1243	B304	Yes

(continued)

Name	Isocode	Glottocode	Guthrie	Data coverage
Poke	pof	poke1238	C53	No
Polri	pmm	pomo1271	A92	Yes
Punu	puu	punu1239	B43	Yes
Sakata	skt	saka1287	C34	Yes
Sake	sak	sake1247	B251	Yes
Sama (Angola)	smd	sama1300	H22	No
San Salvador Kongo	kwy	sans1272	H16a	Yes
Sangu (Gabon)	snq	sang1333	B42	Yes
Seki	syi	seki1238	B21	Yes
Sengele	szg	seng1278	C33	Yes
Shiwa	–	shiw1234	A803	Yes
Sighu	sxe	sigh1238	B202	Yes
Simba	sbw	simb1254	B302	Yes
Sira	swj	sira1266	B41	Yes
So (Cameroon)	sox	soca1235	A82	Yes
So (DRC)	soc	sode1235	C52	Yes
Sonde	shc	sond1250	H321	No
Songo	soo	song1299	B85	No
Songomeno	soe	song1305	C82	Yes
Songoora	sod	song1300	D24	Yes
Suku	sub	suku1259	H32	Yes
Suundi	sdj	suun1239	H131	Yes
Tchitchege	tck	tchi1245	B701	No
Teke-Ebo	ebo	teke1278	B74b	Yes
Teke-Fuumu	ifm	teke1274	B77b	Yes
Teke-Kukuya	kkw	teke1280	B77a	No
Teke-Laali	lli	teke1277	B73	No
Teke-Tege	teg	teke1275	B71	Yes
Teke-Tsaayi	tyi	teke1281	B73	No
Teke-Tyee	tyx	teke1276	B73	No
Tembo (Motembo)	tmv	temb1272	C37	Yes
Tetela	tll	tete1250	C71	Yes
Tibea	ngy	tibe1274	A54	Yes
Tiene	tii	tien1242	B81	Yes
Tsaangi	tsa	tsaa1242	B53	Yes
Tsogo	tsv	tsog1243	B31	Yes
Tuki	bag	tuki1240	A601	Yes
Tunen	tvu	tune1261	A44	Yes
Tuotomb	ttf	tuot1238	A461	No
Ukhwejo	ukh	ukhw1241	A802	Yes
Vanuma	vau	vanu1242	D331	No
Vili	vif	vili1238	H12	Yes
Vili of Ngounie	–	vili1239	B503	No
Viya	gev	eviy1235	B301	Yes

(continued)

Name	Isocode	Glottocode	Guthrie	Data coverage
Vumbu	vum	vumb1238	B403	No
Wandji	wdd	wand1266	B501	No
Wongo	won	wong1247	C85	No
Wumboko	bqm	wumb1241	A22	No
Wumbvu	wum	wumb1242	B24	Yes
Yaka (CAR)	axk	yaka1272	C104	No
Yaka (Congo)	iyx	yaka1274	B73	Yes
Yaka (DRC)	yaf	yaka1269	H31	No
Yambeta	yat	yamb1252	A462	No
Yangben	yav	yang1293	A62	No
Yansi	yns	yans1239	B85	Yes
Yasa	yko	yasa1242	A33a	Yes
Yela	yel	yela1238	C74	Yes
Yombe	yom	yomb1244	H16c	Yes
Zamba	–	zamb1245	C16	Yes
Zimba	zmb	zimb1251	D26	Yes

Appendix C: Additional visualizations

This Appendix contains several figures and/or analyses that support the points raised in the main text in various ways.

We start with Figure 9, which reports on further analyses we conducted on our data set concerning the proposed four-way typology discussed in Section 4.2. While the scatterplot in Figure 3 is clearly illustrative of the patterning of the data, it displays a lot of variation with regard to the number of targets that receive either type of marking. An alternative is to conduct multiple correspondence analysis (MCA), which is used for dimensionality reduction of potentially correlated categorical variables (see Section 3). We conducted MCA on the answers to all the binary questions included in our questionnaire, that is both the questions on syntactic and animacy-based agreement for specific targets, and the set of additional questions which concludes the coding sheet (see Appendix A: Coding model). The results are presented in Figure 9, which reports on the patterning of the two first dimensions of the MCA. These two dimensions together capture half of the variability in the dataset; each following dimension explains a lower and lower proportion of the data.

The results of the MCA analysis depicted in Figure 9 suggest two main clusters of languages, a dense cluster to the center-left and a more lose spread of datapoints

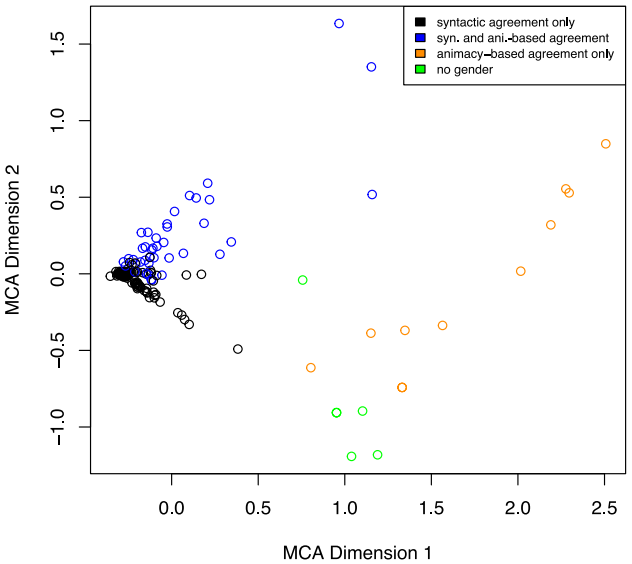

Figure 9: First two dimensions from multiple correspondence analysis (MCA) on the entire questionnaire including the additional questions. The first dimension (*x*-axis) captures 38% of the variance, the second dimension (*y*-axis) captures 12%.

to the center-right of the typological space delimited by the two first dimensions. When projecting the color-coding of the four-way typology we proposed above onto the data points, these clusters become strikingly meaningful: the center-left block corresponds to languages with only syntactic agreement or a combination of syntactic and animacy-based agreemenr. These languages are way more similar to each other than the two other types, that is, languages with only animacy-based gender and languages with no gender, which are scattered throughout the remainder of the space. The second MCA dimension (*y*-axis) distinguishes between languages with only syntactic agreement (in black, negative loading on Dimension 2) and a combination of syntactic and animacy-based agreement (in blue, positive loading on Dimension 2). The MCA thus aligns with the patterns illustrated in Figure 3 in suggesting that the systems of gender marking attested in our sample can be represented and summarized through a four-way classification of types of language structures.

Figure 10 is an alternative to Figure 5. Agreement target types are ordered here by ratio of presence/absence, rather than by frequency of presence. Figure 11 is an alternative visualization to the polygon map presented in Figure 7. Figure 12 helps identify the place of NWB branches within the larger Bantu context.

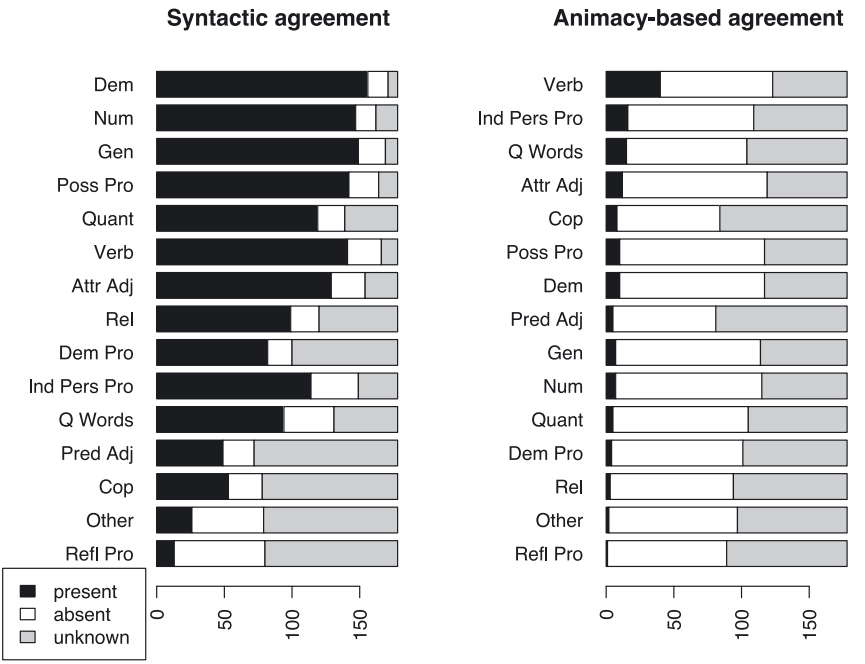

Figure 10: Distribution of syntactic and animacy-based agreement for all targets; ordered by ratio of presence/absence.

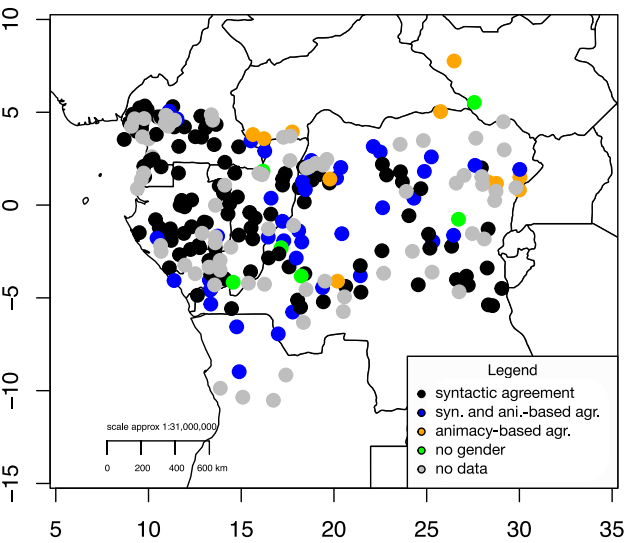

Figure 11: Distribution of types across the sampled area.

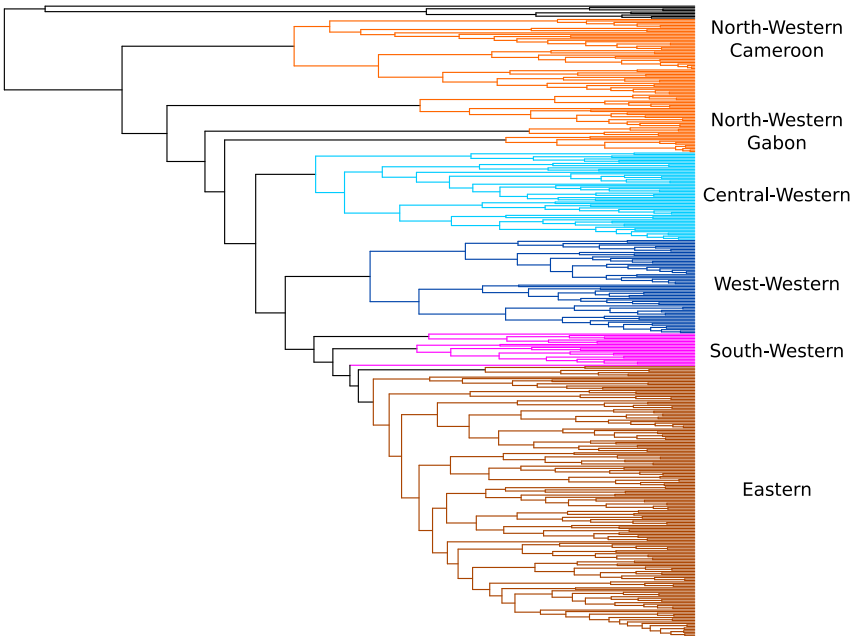

Figure 12: The consensus tree from Grollemund et al. (2015), based on their Figure 1 and annotated with the subgroup names Grollemund et al. (2015) provide.

References

Atindogbe, Gratien. 2013. A grammatical sketch of Mòkpè (Bakweri), Bantu A20. *African Study Monographs: Supplementary Issue* 45. 1–163.

Audring, Jenny. 2019. Canonical, complex, complicated? In Francesca Di Garbo, Bruno Olsson & Bernhad Wälchli (eds.), *Grammatical gender and linguistic complexity*: General issues and areal and language-specific studies, vol. 1, 15–52. Berlin: Language Science Press.

Augustin, MaryAnne. 2010. *Selected features of syntax and information structure in Lika (Bantu D.20)*. Dallas, TX: Graduate Institute of Applied Linguistics MA thesis.

Bearth, Thomas. 2003. Syntax. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 121–142. New York: Routledge.

Beaulieu, Jeremy, Brian C. O'Meara & M. J. Donoghue. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: The evolution of plant habit in campanulid angiosperms. *Systematic Biology* 62. 725–737. https://doi.org/10.1093/sysbio/syt034.

Belliard, François. 2007. *Parlons kwàkùm: langue bantu de l'est Cameroun*. Paris: L'Harmattan.

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

Bostoen, Koen. 2019. Reconstructing Proto-Bantu. In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 2nd edn., 308–334. London: Routledge.

Bostoen, Koen & Hilde Gunnink. Forthcoming. The impact of autochthonous languages on Bantu language variation: A comparative view on Southern and Central Africa. In Salikoko Mufwene & Anna Maria Escobar (eds.), *Cambridge handbook of language contact*. Cambridge: Cambridge University Press.

Boyd, Raymond. 1989. Adamawa-Ubangi. In John Bendor-Samuel & Rhonda L. Hartell (eds.), *The Niger-Congo Languages*, 178–215. Lanham, MD: University Press of America.

Buchanan, Deborah L. 1996/1997. The Munukutuba noun class system. *The Journal of West African Languages* 26(2). 71–86.

Contini-Morava, Ellen. 2008. Human relationship terms, discourse prominence, and asymmetrical animacy in Swahili. *Journal of African Languages and Linguistics* 29(2). 127171.

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

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

Corbett, Greville. 2000. *Number*. Cambridge: Cambridge University Press.

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

Corbett, Greville. 2013a. Number of genders. In Matthew Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. http://wals.info/chapter/30 (accessed 10 April 2021).

Corbett, Greville. 2013b. Sex-based and non-sex-based gender systems. In Matthew Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. http://wals.info/chapter/31 (accessed 10 April 2021).

Corbett, Greville. 2013c. Systems of gender assignment. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. http://wals.info/chapter/32 (accessed 10 April 2021).

Crane, Thera Marie, Larry M. Hyman & Simon Nsielanga Tukumu. 2011. *A grammar of Nzadi [B865]: A Bantu language of the Democratic Republic of Congo*. Berkeley, CA: University of Caifornia Press.

Dahl, Östen. 2000. Animacy and the notion of semantic gender. In Barbara Unterbeck, Matti Rissanen, Tettu Nevalainen & Mirja Saari (eds.), *Gender in grammar and cognition*, 99–115. Berlin & Boston: De Gruyter Mouton.

Dahl, Östen & Kari Fraurud. 1996. Animacy in grammar and discourse. In Thorstein Fretheim & Jeanette K. Gundel (eds.), *Reference and referent accessibility*, 47–64. Amsterdam & Philadelphia: John Benjamins.

de Wit, Gerrit. 2015. *Liko phonology and grammar: A Bantu language of the Democratic Republic of the Congo*, 597. Leiden: Rijksuniversiteit te Leiden dissertation.

Di Garbo, Francesca. 2020. The complexity of grammatical gender and language ecology. In Arkadiev Peter & Francesco Gardani (eds.), *The complexities of morphology*, 193–229. Oxford: Oxford University Press.

Di Garbo, Francesca & Yvonne Agbetsoamedo. 2018. Non-canonical gender in African languages: A typological survey of interactions between gender and number, and gender and evaluative morphology. In Sebastian Fedden, Jenny Audring & Greville Corbett (eds.), *Non-canonical gender systems*, 176–210. Oxford: Oxford University Press.

Dimmendaal, Gerrit. 2000. Number marking and noun categorization in Nilo-Saharan languages. *Anthropological Linguistics* 42. 214–261.

Downing, Laura & Lutz Marten. 2019. Clausal morphosyntax and information structure. In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 2nd edn., 270–307. London: Routledge.

Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson & Russel D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473. 79–82.

Enger, Hans-Olav & Tore Nesset. 2011. Constraints on diachronic development: The animacy hierarchy and the relevance constraint. *STUF Language Typology and Universals* 64. 193–212.

Ernst, Urs. 1992. *Esquisse grammaticale du kako*. Yaoundé: Société internationale de linguistique.

Faraclas, Nicholas. 1986. Cross River as a model for the evolution of Benue-Congo nominal class/concord systems. *Studies in African Linguistics* 17(1). 39–54.

Fiedler, Ines, Tom Güldemann & Benedikt Winkhart. 2021. The two concurrent systems of Mba. *STUF Language Typology and Universals* 74. 303–325.

Fritz, Susanne A. & Andy Purvis. 2010. Selectivity in mammalian extinction risk and threat types: A new measure of phylogenetic signal strength in binary traits. *Conservation Biology* 24. 1042–1051.

Gérard, [R. P.]. 1924. *La langue lebéo, grammaire et vocabulaire* (Bibliotèque-Congo 13). Brussels: A. Vromant & Co.

Good, Jeff. 2012. How to become a "Kwa" noun. *Morphology* 22. 293–335.

Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 73–113. Cambridge, MA: MIT Press.

Grollemund, Rebecca, Brandford Simon, Koen Bostoen, Andrew Meade, Chris Venditti & Mark Pagel. 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences of the United States of America* 112. 13296–13301.

Grollemund, Rebecca, Jean-Marie Hombert & Branford Simon. 2018. A phylogenetic study of North-Western Bantu and South Bantoid languages. In Rajend Meshtrie & David Bradley (eds.), *The dynamics of language: Plenary and focus lectures from the 20th International Congress of Linguists. Cape Town*, Cape Town, July 2018, 118–132. UCT Press.

Güldemann, Tom & Ines Fiedler. 2019. Niger-Congo "noun classes" conflate gender with deriflection. In Francesca Di Garbo, Bruno Olsson & Wälchli Bernard (eds.), *Grammatical gender and linguistic complexity, volume I: General issues and specific studies*, 95–145. Berlin: Language Science Press.

Guthrie, Malcolm. 1948. *The classification of the Bantu languages*. Oxford: Oxford University Press.

Hammarström, Harald, Robert Forkel & Martin Haspelmath (eds.). 2019. *Glottolog 3.4*. Jena: Max Planck Institute for the Science of Human History.

Harries, Lyndon. 1958. Kumu, a sub-Bantu language. *Kongo-Overzee* 24. 265–296.

Hyman, Larry M. 1980. Esquisse des classes nominales en tuki. In Larry M. Hyman (ed.), *Noun classes in the Grassfield Bantu borderland (Southern California Occasional Papers in Linguistics 8)*, 27–36. Los Angeles, CA: Dept. of Linguistics, University of Southern California.

Hyman, Larry M. 2003. Basaá. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 257–282. London & New York: Routledge.

Idiatov, Dmitry. 2007. *A typology of non-selective interrogatives*. Antwerp: University of Antwerp dissertation.

Igartua, Iván & Ekaitz Santazilia. 2018. How animacy and natural gender constrain morphological complexity: Evidence from diachrony. *Open Linguistics* 4(1). 438–452.

Iorio, David. 2011. The noun phrase in Kibembe (D54). *Newcastle Working Papers in Linguistics* 17. 46–65.

Isaac, Kendall M. 2014. *Noun classes in Lefa* (ALACAM581). Yaoundé: SIL.
Karatsareas, Petros. 2009. The loss of grammatical gender in Cappadocian Greek. *Transactions of the Philological Society* 107. 196–230.
Karatsareas, Petros. 2014. On the diachrony of gender in Asia Minor Greek: The development of semantic agreement in Pontic. *Language Sciences* 43. 77–101.
Katamba, Francis. 2003. Bantu nominal morphology. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 103–120. London: Routledge.
Koile, Ezequiel, Simon J. Greenhill, Damian E. Blasi, Remco Bouckaert, Tom Güldemann, Patrick Roberts & Russell D. Gray. Under review. Phylogeographic analysis of the Bantu expansion supports a rainforest route.
Kutsch Lojenga, Constance. 2003. Bila (D 32). In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 450–474. London: Routledge.
Lê, Sébastien, Julie Josse & François Husson. 2008. Factominer: an R package for multivariate analysis. *Journal of Statistical Software* 25(1). 1–18.
Leitch, Myles. 2003. Babole (C 101). In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 392–421. London: Routledge.
Luraghi, Silvia. 2011. The origin of the Proto-Indo-European gender system: Typological considerations. *Folia Linguistica* 45. 435–464.
Maho, Jouni. 1999. *A comparative study of Bantu noun classes. (Acta universitatis gothoburgensia)*. Göteborg: Orientalia et Africana Gothoburgensia dissertation.
Marchese, Lynell. 1988. Noun classes and agreement systems in Kru: A historical approach. In Michael Barlow & Charles A Ferguson (eds.), *Agreement in natural languages: Approaches, theory, descriptions*, 323–341. Stanford: CA: CSLI.
Matasovic, Ranko. 2004. *Gender in Indo-European*. Heidelberg: Winter.
Meeussen, Achille E. 1967. Bantu grammatical reconstructions. *Africana Linguistica* 3. 79121.
Meeuwis, Michael. 2013. Lingala. In Susanne Michaelis Philipe, Martin Haspelmath Maurer & 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.
Merlevede, Andrea. 1995. *Een schets van de fonologie en morfologie van het Bondei gevolgd door een Bondei-Engels en Engels-Bondei woordenlijst*. University of Leiden MA thesis.
Mfoutou, Jean-Alexis. 2009. *Grammaire et lexique munukutuba: Congo-Brazzaville, République démocratique du Congo, Angola*. Paris: LHarmattan.
Motingea Mangulu, André. 2005. *Leboale et lebaati: langues bantoues du plateau des Uélé, Afrique Centrale* (ILCAA Language Monograph Series 2). Tokyo: Institute for the Study of Languages, Cultures of Asia and Africa (ILCAA).
Motingea Mangulu, André. 2008. *Aspects du bongili de la Sangha-Likouala, suivis de l'esquisse du parler énga de Mampoko, Lulonga* (ILCAA Language Monograph Series 4). Institute for the Study of Languages, Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies.
Motingea Mangulu, André. 2010. *Aspects des parlers minoritaires des Lacs Tumba et Inongo: Contribution a l'histoire de contact des langues dans le bassin central congolais* (ILCAA Language Monograph Series 5). Tokyo: Institute for the Study of Languages, Cultures of Asia and Africa (ILCAA).
Mufwene, Salikoko. 1997. Kitúba. In Sarah Grey Thomason (ed.), *Contact languages: A wider perspective* (Creole language library 17), 173–208. Amsterdam & Philadelphia: John Benjamins.

Nurse, Derek & Gérard Philippson. 2003a. Introduction. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 1–12. London: Routledge.

Nurse, Derek & Gérard Philippson. 2003b. Towards a historical classification of the Bantu languages. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 164–182. London: Routledge.

Ollomo Ella, Régis. 2013. *Description linguistique du shiwa, langue bantu du Gabon: Phonologie, morphologie, syntaxe, lexique*. Paris: Université de la Sorbonne Nouvelle (Paris 3) dissertation.

Orme, David, Robert P. Freckleton, Thomas Gavin, Thomas Petzoldt, Susanne Fritz, Nick Isaac & Will Pearse. 2013. The caper package: Comparative analysis of phylogenetics and evolution in R. *R package version* 5(2). 1–36.

Pagel, Mark. 1994. Detecting correlated evolution on phylogenies: a general method for the comparatuve analysis of discrete characters. *Proceedings of the Royal Society (B)* 255. 37–45.

Philippson, Gérard & Rebecca Grollemund. 2019. Classifying Bantu languages. In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 2nd edn., 335–354. New York: Routledge.

R Core Team. 2018. *R: A Language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Reeder, Jedene. 1998. *Pagibete, A northern Bantu borderlands language: A grammatical sketch (Congo)*. Arlington, TX: University of Texas at Arlington MA thesis.

Revell, Liam J. 2012. Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3. 217–223.

Richardson, Irvine. 1957. *Linguistic survey of the Northern Bantu Borderland*. London: Oxford University Press.

Santandrea, Stefano. 1963. Short notes on the Bodo, Huma and Kare languages. *Sudan Notes and Records* 44. 82–99.

Schadeberg, Tilo. 2003. Historical linguistics. In Gérard Philippson & Derek Nurse (eds.), *The Bantu languages*, 143–163. New York: Routledge.

Seifart, Frank. 2018. The semantic reduction of the noun universe and the diachrony of nominal classification. In William McGregor & Søren Wichmann (eds.), *The diachrony of classification systems*, 9–32. Amsterdam & Philadelphia: John Benjamins.

Smith-Stark, Cedric. 1974. The plurality split. *Chicago Linguistic Society* 10. 657–671.

Stappers, Leo. 1955. Schets van het Budya. *Kongo-Overzee* 21. 97–143.

Stucky, Suzanne U. 1978. How a noun system may be lost: Evidence from Kituba (lingua franca Kikongo). *Studies in the Linguistic Sciences* 8(1). 216–233.

Susa, Dz'ba Dheli. 1972. *Esquisse grammaticale du bira*. Lubumbashi: Université Nationale du Zaïre (UNAZA) MA thesis.

Thomas, John Paul. 1994. Bantu noun-class reflexes in Komo. *Africana Linguistica* 142. 177–195.

Thornell, Christina. 2010. Morphology of plant names in the Mpiemo language. In Karsten Legère, Christina Thornell, Bernd Heine & Wilhelm J. G. Möhlig (eds.), *Bantu languages: Analyses, description and theory*, 249–270. Cologne: Rüdiger Köppe Verlag.

Thornell, Christina. 2012. Simplification of the nominal class system in Central African Bantu Bendo [bndo]. *Paper presented at the 7th World Congress of African Linguistics (WOCAL)*.

Van de Velde, Mark. 2008. *A grammar of Eton*. Berlin & New York: Mouton de Gruyter.

Van de Velde, Mark. 2019. Nominal morphology and syntax. In Koen Bostoen & Mark van de Velde (eds.), *The Bantu languages*, 2nd edn., 237–269. New York: Routledge.

Van de Velde, Mark. 2021. The AMAR mechanism: Nominal expressions in the Bantu languages are shaped by apposition and reintegration. *Linguistics* 60(3). 1–33. Published online ahead of print.

Van Epps, Briana. 2019. *Sociolinguistic, comparative and historical perspective on Scandinavian gender: With focus on Jamtlandic*. Lund: Lund University dissertation.

Verkerk, Annemarie & Francesca Di Garbo. 2022. Sociogeographic correlates of typological variation in norhwestern Bantu gender systems. *Language Dynamics and Change*. 1–69. Published online ahead of print 2022. https://doi.org/10.1163/22105832-bja10017.

Vihman, Virve-Anneli, Diane Nelson & Simon Kirby. 2018. Animacy distinctions arise from iterated learning. *Open Linguistics* 4(1). 552–565.

Wald, Benji V. 1975. Animate concord in northeast coastal Bantu: Its linguistic and social implications as a case of grammatical convergence. *Studies in African Linguistics* 6. 267–314.

Wega, Simeu Abraham. 2012. *Grammaire descriptive du pólrò: eléments de phonologie de morphologie et de syntaxe*. Yaounde: Universite de Yaounde I - Department de langues africaines et linguistique dissertation.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/ling-2020-0217).