Olena Shcherbakova, Marc Allassonnière-Tang and
Francesca Di Garbo*

The evolution of gender and number agreement in the noun phrase

https://doi.org/10.1515/lingty-2024-0072
Received September 19, 2024; accepted May 21, 2025; published online August 15, 2025

Abstract: We test the dependency relations between gender and number posited by Greenberg's Universal 36 by focusing on patterns of gender and number agreement in the noun phrase from an evolutionary perspective. To do so, we use data from Grambank, the largest existing database of morphosyntactic structures in the world's languages. Based on data from 1,608 languages worldwide, we use a Reverse Jump Markov Chain Monte Carlo method to investigate the order of emergence of gender and number marking on adjectives, demonstratives and (for a smaller dataset) articles. Globally, our findings support Greenberg's idea that gender marking hinges on number marking. In addition, they show that both adjectives and demonstratives play a special role in the development of noun-phrase internal agreement, in that under different evolutionary scenarios the occurrence of gender and number agreement on adjectives or demonstratives may favor the spreading of agreement to other target types. We compare these results with family-specific patterns of language change and further discuss their relevance to the general understanding of nominal morphosyntax.

Keywords: gender; number; agreement; phylogenetics

1 Introduction

By proposing the implicational relation between gender and number, Greenberg's Universal 36 ("If a language has the category of gender it always has the category

***Corresponding author: Francesca Di Garbo, [fran'tʃeska di 'garbo]**, Aix-Marseille Univ. – CNRS LPL, UMR 7309, Aix-en-Provence, France, E-mail: francesca.di-garbo@univ.fr
Olena Shcherbakova [o'lɛnɑ ʃtʃɛr'bɑkɔvɑ], Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, E-mail: olena_shcherbakova@eva.mpg.de
Marc Allassonnière-Tang [mɑʁk alasɔŋjɛʁ tɑ̃ŋ], Lab Eco-Anthropology UMR 7206, CNRS/MNHN/ University Paris City, Paris, France, E-mail: marc.allassonniere-tang@mnhn.fr. https://orcid.org/0000-0002-9057-642X

number", Greenberg 1963) has inspired a number of typological studies. This paper takes a further step in this debate. It explores the interdependence between gender and number in the languages of the world by examining evolutionary relationships between these categories in terms of the order of emergence of gender and number agreement.

Gender and number are two features of nominal morphosyntax which contribute to reference construal and reference tracking, that is, to the linguistic representation of real world entities. Gender is a nominal classification strategy which, in the languages that possess it, functions as an inherent lexical property of nouns (Corbett 1991). Gender distinctions distribute the nominal lexicon of a language into two or several classes whose semantic motivation varies a great deal in the languages of the world, while generally revolving around such notions as animacy, sex, size and shape. Number is a feature of inherent nominal inflection. Number distinctions serve the purpose of representing nouns and noun phrases as denoting one or several instances of an entity (Corbett 2000). Plurality is the most frequently attested type of number value, and the one that is also most likely to be obligatorily coded across the languages of the world (Corbett 2000; Greenberg 1963). The existence of gender and number in the languages of the world is often connected. For instance, if both present, gender and number are generally encoded through cumulative morphemes (Di Garbo 2014: and references therein).

Morphosyntactically, the key difference between gender and number is that the presence of gender agreement is the prerequisite to establish that a grammatical gender system exists in a given language, whereas the presence of number marking can be restricted to nouns only, or to the noun phrase level (Corbett 2013). In general, number tends to be marked on nouns more frequently than gender does. For instance, in the languages of Africa, out of a sample of 84 languages with gender, 15 of them do not have gender marking on nouns, while all display some form of number marking on nouns or at the noun phrase level (Di Garbo 2014: 105). Finally, on a global scale, number systems are more frequent than gender systems. For instance, taking the World Atlas of Language Structures (WALS) database as a reference, out of a sample of 257 languages coded for gender, 145 display no grammatical gender systems. In contrast, of a sample of 1,066 languages coded for the marking of nominal plurality, only 98 show no instance of nominal plural marking (Corbett 2013; Haspelmath 2013).

As mentioned above, the categories of gender and number have been suggested to be interdependent in the languages of the world. This generalization is reflected in Greenberg's (1963: 58) Universal 36: "If a language has the category of

gender, it always has the category of number". This universal establishes an asymmetry about the existence of gender and number in the languages of the world, such that the presence of grammatical gender (as a general morphosyntactic category) always hinges on the existence of some type of number system (also as a general morphosyntactic category). In other words, languages possessing gender and lacking number are expected to be rare, whereas languages with number only or both gender and number are more common. From a language evolution perspective, this would suggest that before the gender category can emerge, a language should have developed number first. Similarly, other works propose the dependence of gender on the availability of number or other morphosyntactic categories in a language. Specifically, the emergence of gender has been described as "parasitic on other category types, notably number and case" (Wälchli et al. 2019: 267) and the presence of person-number agreement (Nichols 1992: 142). In other words, if gender emerges in a language, this happens after the number category (often in combination with person or case) has already been grammaticalized. Furthermore, the development of the gender category has been suggested to influence the morphosyntactic complexity of number marking. For example, it has been observed that African languages with both number and gender categories typically mark these categories on a variety of targets, whereas in the absence of the gender category, number marking tends to be restricted to nouns or to the phrasal level (Di Garbo 2014: based on a sample of 100 African languages).

The validity of Greenberg's Universal 36 has been recently tested by Verkerk et al. (forthc.), along with many other typological universals. The study uses the Grambank dataset (Skirgård et al. 2023) and the global EDGE tree (Bouckaert et al. 2022) as the reference classification for phylogenetic analysis. In line with Greenberg's Universal 36, they find strong evidence supporting the correlated evolution between gender and number worldwide, which confirms that the two features are interdependent. However, the study does not go into detail on whether gender tends to emerge only after number is already present. Furthermore, the co-evolutionary relationship holds for gender and number categories overall, that is, not taking into account where exactly these categories are marked. In other words, a language possesses gender or number if the respective marking is coded as present in general or for at least one word class the information for which is available in Grambank (nouns, numerals, articles, adjectives, demonstratives, etc.). Thus, while the results of this study yield cross-linguistic support for the interdependence of number and gender, it is unclear if gender indeed hinges on number as proposed by Greenberg. Thus, two open questions remain.

First, as mentioned above, the "gender hinges on number" hypothesis has been examined only from the broad perspective of the overall existence of the two categories rather than in terms of their actual morphosyntactic manifestations. Given that the existence of grammatical gender is, by definition, tied to agreement, it would be important to establish tendencies in the evolution of gender and number agreement across different targets. With respect to the noun phrase domain, for instance, does number agreement precede the development of gender agreement on adjective and demonstratives, or is the opposite trend attested for these target types? Second, since the study by Verkerk et al. (forthc.) has a global focus, it remains unclear whether and to which extent individual language families might follow these global tendencies. Our study aims to fill both of these gaps and identify the order of emergence of gender and number globally, and in the context of specific language families.

Additionally, focusing on the evolution of gender and number agreement across different types of targets within the noun phrase positions us to answer the question about which type of target shows agreement first. The idea that agreement on demonstratives precedes agreement on adjectives is stated in the form of an implicational generalization by Moravcsik (1997: 317): "in more than one language, and possibly in all, if the adnominal adjective agrees with the noun, so does the adnominal demonstrative". However, to the best of our knowledge, this claim has never been tested quantitatively on a large sample of languages.

In sum, while the implicational relationship between gender and number as morphosyntactic categories has been studied both qualitatively and quantitatively, the exact evolutionary dynamics that inform this relationship in the domain of gender and number agreement have remained overlooked. Here, we focus on the order in which gender and number agreement develops across different targets in the nominal phrase. We test if their evolutionary dynamics on the global tree and in individual language families mirror Greenberg's "gender hinges on number" scenario about the development of these categories; that is, the fact that gender agreement tends to emerge in languages that already possess number agreement. Furthermore, our approach allows us to shed light on whether adjectives develop gender and/or number agreement before respective agreement emerges on demonstratives.

2 Data and method

To extract typological information, we use Grambank (Skirgård et al. 2023), the global database covering a variety of grammatical phenomena in 195 features, including gender and number agreement on various targets. We obtain the features indicating whether demonstratives, adjectives, and articles (definite and/or indefinite) agree

with the nouns in number (GB184, GB185, and GB186) and gender/noun class (GB170, GB171, and GB172). We restrict our analyses to Grambank features related to agreement on those three targets because data on other targets was either completely unavailable or limited to only one type of agreement. For example, Grambank has a feature on gender agreement in numerals, but lacks one for number agreement in numerals. Similarly, there are no features allowing tracking person agreement. In this process, several methodological decisions are made. First, throughout the paper, we use the term "adjectives" for convenience, even though not all languages have a distinct word class of adjectives (Paul 2010). The relevant Grambank features focus on whether agreement is available in "property words" (equivalent to "adjectives" only in some linguistic settings) that encode properties of nominal expressions. Second, by including all three targets in the nominal domain, we attempt to get the full picture of the evolution of number and gender agreement in the noun phrase. This could additionally shed light on how agreement develops in articles and demonstratives or whether articles are more likely to agree in number and/or gender if demonstratives do as well. However, definite articles are absent in many languages (1,374/2,198 languages in Grambank), thus, including only languages that have articles would restrict the scope to specific language families, which might create a strong sample bias. Therefore, we conduct two analyses.

The first analysis considers only languages that have articles (n = 662) and offers the overview of the evolution of number and gender across all three targets: demonstratives, adjectives, and articles. Here we exclude languages without articles to ensure that they are not assigned the same value of zero as languages with articles but no number/gender marking. The second analysis has fewer features (number/gender marking on demonstratives and adjectives only) but a larger sample of languages (n = 1,608). As an example of how the data is encoded, some languages like Italian mark both gender and number on all three targets, as shown in example (1).

(1) Gender and number marking on definite articles and demonstratives in Italian (Indo-European, personal knowledge)

(a) *il/questo* *muro* *bianco*
the.MASC.SG/this.MASC wall.MASC.SG white.MASC.SG
'The/this white wall'

(b) *la/questa* *casa* *bianca*
the.FEM.SG/this.FEM house.FEM.SG white.FEM.SG
'The/this white house'

(c) *i/questi* *muri* *bianchi*
the.MASC.PL/this.MASC wall.MASC.PL white.MASC.PL
'The/these white walls'

(d) *le/queste* *case* *bianche*
the.FEM.PL/this.FEM house.FEM.PL white.FEM.PL
'The/these white houses'

French behaves in a similar way with the exception that gender agreement in articles (*les grandes maisons* 'the.PL big.FEM.PL houses.PL') and demonstratives *(ces grandes maisons* 'these.PL big.FEM.PL houses'.PL)) is neutralized in the plural, whereas number agreement is consistent across all targets. In English, only demonstratives agree with nouns in number (*this flower/these flowers*) while articles and adjectives remain unchanged. As English nouns are not categorized into genders, none of the targets have gender agreement. Yet, in some languages (e.g. Māngarongaro (Austronesian), gender and number agreement is absent across all three targets.

Languages marking 1) only number, 2) both number and gender, or 3) neither number nor gender would satisfy Greenberg's "gender hinges on number" prediction. However, we encounter various instances of languages having gender agreement and lacking number agreement in specific targets. For instance, in the isolate Tunica, adjectives can take gender marking, but do not agree with nouns in number. In Tunni (Afro-Asiatic) and Marshallese (Austronesian), number (but not gender) is marked on adjectives in line with Greenberg's claims, but demonstratives and adjectives show the opposite pattern of agreeing with nouns in gender, but not in number. These different scenarios are summarized and compared in Table 1.

We use phylogenetic comparative methods to infer the diachronic interaction between the variables of our dataset. Such methods allow us to address the non-independence of features from evolutionary processes (Galton's problem, Mace and Holden 2005; Macklin-Cordes and Round 2022). We use a global tree of 6,635 languages (Bouckaert et al. 2022). This tree incorporates information on language classification within families from previously published phylogenies and Glottolog (Hammarström et al. 2021). It infers the relationships between language families and isolates based on geographic locations of the languages. The time calibrations result from a wide range of archeological, genetic, and linguistic evidence as well as published phylogenies. We use the entire sample of 902 posterior trees of this global tree

Table 1: Illustration of the features and feature values considered in the study.

Languages	Adjectives		Articles		Demonstratives	
	Number	Gender	Number	Gender	Number	Gender
Italian, French	1	1	1	1	1	1
Māngarongaro	0	0	0	0	0	0
Tunica, Tunni	1	0	0	1	0	1

to control for phylogenetic non-independence between languages. Each of those trees represents the evolutionary processes underlying the divergence of languages and thus allows us to appropriately capture uncertainty in our results associated with possible evolutionary relationships between languages.

As mentioned above, we conduct two analyses using two different samples. First, we focus on number and gender marking on adjectives, demonstratives, and articles. Since we exclude languages that do not possess articles or do not have information on whether or not articles agree with gender or number, our global sample for this analysis is restricted to 662 languages. We prune the world tree to keep languages for which we have data for the six Grambank features (gender and number marking on adjectives, demonstratives, and articles). We then conduct phylogenetic analyses to assess the diachronic interaction between these six variables. For this analysis, we focus on the global tendencies observable in the languages that have articles and do not conduct analyses for specific language families. This is done to prevent making inferences about language families that show variation in the distribution of articles. For instance, in the Indo-European family, only some branches (Germanic and Romance) tend to have articles, whereas articles are mainly absent in other branches (Slavic and Indo-Aryan), which would need to be excluded from such an analysis. Figure 1 shows the geographic distribution of the data set for this analysis.

Second, we consider the diachronic interaction between number and gender marking in adjectives and demonstratives. This analysis allows for a larger coverage of languages and language families. Our global sample (n = 1,608) includes languages found on the global tree and in Grambank with available data on presence/absence of number and gender marking on adjectives and demonstratives. Apart from assessing the diachronic interaction of number and gender on the world tree, we also consider language families with sufficient data in typological variation of gender/number marking and available phylogenetic trees, i.e. the Indo-European language family

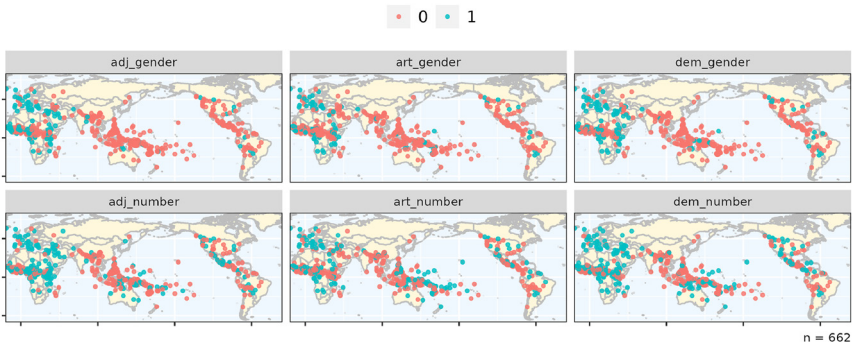

Figure 1: The geographic distribution of languages having data available for the presence/absence of number/gender marking on adjectives, articles, and demonstratives.

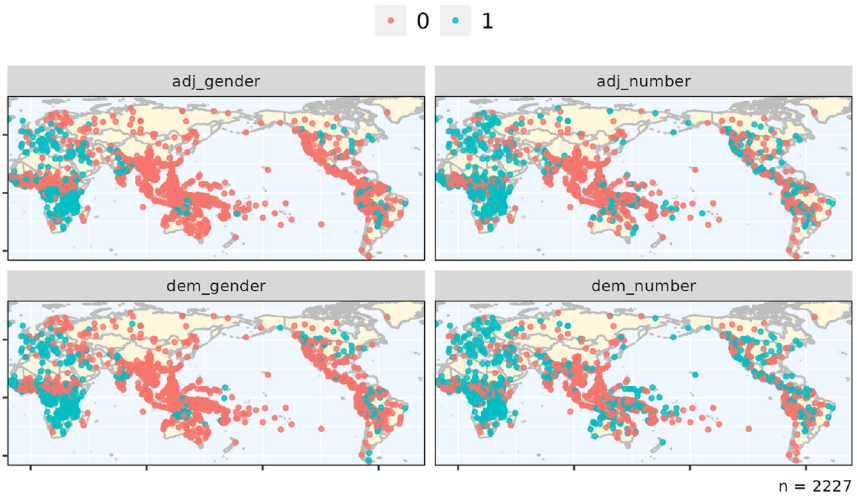

Figure 2: The geographic distribution of languages having data available for the presence/absence of number/gender marking on adjectives and demonstratives.

(42 languages; Bouckaert et al. 2012), the Dravidian language family (15 languages; Kolipakam et al. 2018), and the Bantu branch of the Atlantic-Congo language family (118 languages; Grollemund et al. 2015). Figure 2 shows the geographic distribution of the data set for this second analysis on the global tree.

To measure the diachronic interaction between the variables, we use a Reverse Jump Markov Chain Monte Carlo (RJ-MCMC) within *BayesTraits V4* software (Green 1995; Gowri-Shankar and Rattray 2007; Pagel et al. 2004). We use this method to test the probability of change and transition rates related to gender and number marking in the noun phrase. The RJ-MCMC is based on a Continuous Time Markov Chain process, which considers scenarios of reversed change between different states of a variable. For example, the algorithm scores the probability that gender marking on adjectives is lost, acquired, re-acquired and re-lost. The RJ-MCMC method thus allows us to infer the correlated evolution between the features included in our dataset. In other words, this approach lets us detect evolutionary relationships between two binary linguistic variables where presence/absence of one variable influences the probability of change in another variable and vice versa (Pagel 1994: 40). We set a Reverse-Jump hyper-prior for all transition rates to a gamma prior with both parameters equal to 1 ("RevJump gamma 1 1"). We can infer whether gender marking or number marking on adjectives is more likely to be acquired (or lost) first and which state is the most stable. To operationalize this, we consider the pairwise correlated evolution between the variables of the data set. Each pair of variables can

have four states: 00, 01, 10, and 11. For example, if the pair under consideration is gender marking on the adjective and number marking on the adjective, 00 would indicate a state where neither gender marking nor number marking is found on the adjective. 11 would represent the state where both gender and number are marked on the adjective. 10 would refer to the situation in which gender is marked on the adjective and not number, while 01 would mean that gender is not marked on the adjective but number is marked. The transition rates can be read as follows based on the same example. If neither gender nor number is marked on the adjective, the higher transition rate from 00 to 10 compared to 00 to 01 suggests that adjectives are more likely to develop gender marking than number marking. Conversely, equal transition rates from 00 to 10 and from 00 to 01 indicate that gender marking and number marking are equally likely to develop on the adjective starting from a state where neither is marked.

For each pair of variables, two models are built. One model considers that the variables are independent and the other model includes the dependence between the variables. For each pair of models, we calculated Bayes Factors (Burnham & Anderson 2004) from the marginal likelihoods of both models which we obtained using a stepping stone sampler (Xie et al. 2011) with 100 stones and 1,000 iterations per stone. For each model, 1,000,000 iterations are conducted. The first half (500,000) is discarded as a burn-in and the sampling frequency is every 1,000 iterations, which results in (1,000,000–500,000)/1,000 = 500 iterations per pair of variables. The Bayes Factor is estimated in the following way: 2 x (log marginal likelihood of dependent model – log marginal likelihood of independent model). We interpret Bayes Factors above 2 as positive evidence, above 5 as strong, and above 10 as very strong evidence in support of the dependent model (Raftery 1996). In the current study, we kept the pairwise interactions that had a Bayes Factor higher than 2. The detailed Bayes factors are listed in the Supplementary Materials.

We use the analyses output of transition rates to examine evolutionary pathways of gaining number/gender marking on three targets by focusing on two major questions. Firstly, we aim to establish which of two features is more likely to emerge in the language with both features absent. For example, we can observe the transition rates from Adj(∅, ∅) to Adj(N, ∅) and to Adj(∅, G) and determine if languages without any marking have preference for either number or gender marking. Furthermore, this approach enables us to compare transitions between different targets, for instance, from N(∅, ∅) to N(Adj, ∅) and to N(∅, Dem) to identify on which target number marking tends to develop first.

Secondly, we are interested in how two features in the observed pairs come to co-occur, specifically identifying which feature was gained first and which was gained second. For instance, this allows us to compare the transition rates towards Adj(N, G) from Adj(N, ∅) and from Adj(∅, G) to establish a potential

tendency for adjectives to gain gender when number is present or vice versa. Similarly, such comparison could shed light on whether number marking on both targets N(Adj, Dem) arises in a specific order (first on adjectives, then on demonstratives or vice versa).

We extract the posterior distribution of the transition rates of interest in co-evolving feature pairs (i.e. those where calculated Bayes Factors exceed 2). We calculate the difference between the distributions of two selected transition rates, for instance, of gaining number marking Adj(N, ∅) and gaining gender marking Adj(∅, G) starting from the state where both are absent Adj(∅, ∅). As a result, we obtain the distribution of values representing the differences between these two transition rates, which we then summarize by calculating the medians and 95 % highest-density continuous intervals (as some posterior distributions are multimodal) using *ggdist* package (Kay 2023). It is then possible to differentiate between cases where one of the rates exceeds the other and where both rates are comparable. Specifically, if the 95 % intervals of the different values exclude zero, we interpret it as evidence for one of the rates being substantially faster than the compared rate, suggesting a specific order of gaining two features. Otherwise we have no conclusive evidence suggesting that transition from the initial state Adj(∅, ∅) towards either Adj(N, ∅) or Adj(∅, G) would be dominant. Then, for variables for which the order of gaining was established, we capture the relationships between these variables with the help of directed arrows.

The analyses are conducted using the *BayesTraits* software (Pagel et al. 2004); data processing and visualization are done in *R* (R Core Team 2023) with the following packages: *ape* (Paradis and Schliep 2018), *diagram* (Soetaert 2020), *GGally* (Schloerke et al. 2020), *ggpubr* (Kassambara 2023), *phangorn* (Schliep 2011), *phytools* (Revell 2012), *tidyverse* (Wickham 2017), *ggdist* (Kay 2023), *ggraph* (Pedersen 2022), *igraph* (Csardi and Nepusz 2006), *ggokabeito* (Barrett 2021), *phylopath* (van der Bijl 2018).

3 Results

3.1 Six-feature analysis: languages with articles

3.1.1 Order of developing gender and number marking on the same and different targets

First, we obtain the differences of transition rates and their medians from the global analysis with six variables: number and gender marking on adjectives, demonstratives, and articles. Figure 3(a) shows the global tendency of gender hinging on number: when both categories are unmarked on the same target, number agreement is more likely to develop first. For example, number marking on articles

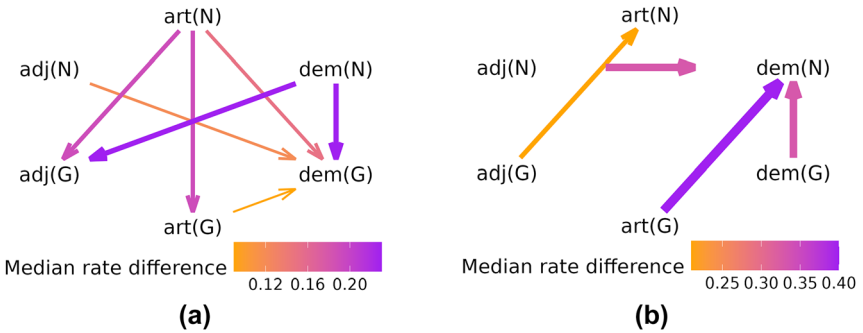

Figure 3: Six features, world tree. The color of the arrows indicates the median difference of transition rates between two features when starting from a state when both features are absent (a) or arriving at a state when both are present (b). The direction of arrows represents the order of gaining the features. The absence of arrows suggests no conclusive evidence in favor of a specific direction between two features.

and demonstratives is more likely to develop before gender marking on these targets. For adjectives, there is no clear tendency in the order of acquiring number and gender marking. Additionally, we observe the tendency of gender hinging on number on distinct targets. For instance, number on articles tends to be acquired before gender on adjectives and demonstratives, number on demonstratives is typically developed before gender on adjectives, and number on adjectives tends to precede gender on demonstratives.

On the other hand, Figure 3(b) reveals the opposite trend in the rate differences: when considering the transition to a state where both variables are marked on the same target, number marking tends to follow gender marking. We find gender marking preceding number marking on the demonstratives, but no clear pattern in the order of developing number and gender marking on articles and adjectives. Furthermore, we observe a tendency whereby number marking follows gender marking across different targets, such as number marking on articles following gender marking on adjectives, and number marking on demonstratives following gender marking on articles.

3.1.2 Cross-target patterns in evolution of gender and number

Transition rates concerning different targets also reveal if number or gender marking tends to first appear on a specific target before it is marked on another target. For instance, we find that in languages without gender marking on either articles or demonstratives, articles gain gender marking before demonstratives

develop gender marking. Another example is the tendency for languages with number marking on adjectives to gain number also on demonstratives.

3.2 Four-feature analysis

3.2.1 Order of developing gender and number marking on the same and different targets

Then, we consider the transition rate differences from the global analysis with four variables: number and gender marking on adjectives and demonstratives. As highlighted in Section 2, considering only adjectives and demonstratives as a locus of marking allows us to rely on a much larger dataset than when including languages that have articles as was done in our analysis with six features. The results of this analysis generally mirror two tendencies captured by the analysis on six features.

Firstly, similar to the results of the analysis on six features, Figure 4(a) illustrates the tendency of gender hinging on number when analyzing transitions from the absence of marking to acquiring either gender or number marking. This pattern of first developing number and then gender holds for both demonstratives and adjectives. Secondly, Figure 4(b) demonstrates that gender marking is developed before number marking in both demonstratives and adjectives before both targets agree in gender and number.

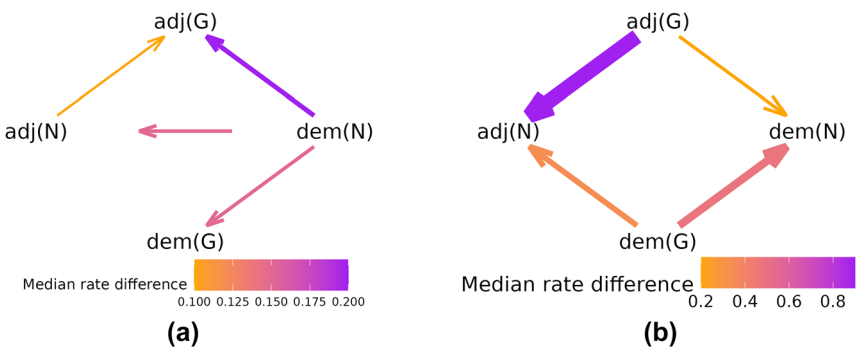

Figure 4: Four features, world tree. The color of the arrows indicates the median difference of transition rates between two features when starting from a state when both features are absent (a) or arriving at a state when both are present (b). The direction of arrows represents the order of gaining the features. The absence of arrows suggests no conclusive evidence in favor of a specific direction between two features.

Overall, this analysis on a larger sample further supports the evolutionary trajectories of agreement development in demonstratives (already evidenced by the four feature analysis) and additionally suggests that agreement in adjectives is gained in the same order. These findings can be summarized as the following two scenarios:

1. When neither number nor gender are marked on the same agreement target, then the development of number marking precedes gender i.e.,

$$\text{Dem}(\varnothing, \varnothing) \text{ and } \text{Adj}(\varnothing, \varnothing)$$

are more likely to transition to

$$\text{Dem}(N, \varnothing) \text{ and } \text{Adj}(N, \varnothing),$$

respectively, rather than to

$$\text{Dem}(\varnothing, G) \text{ and } \text{Adj}(\varnothing, G).$$

2. When both gender and number are marked on the same target, then the preceding stage is the presence of gender marking and the absence of number marking i.e,

$$\text{Dem}(\varnothing, G) \text{ and } \text{Adj}(\varnothing, G),$$

respectively, rather than

$$\text{Dem}(N, \varnothing) \text{ and } \text{Adj}(N, \varnothing)$$

that is more likely to transition to

$$\text{Dem}(N, G) \text{ and } \text{Adj}(N, G).$$

3.2.2 Cross-target patterns in evolution of gender and number

On the larger sample, we observe two trends. When considering the situation in which agreement emerges from a state in which both features are absent (Figure 4a), number marking on demonstratives emerges first to later spread to adjectives. In this scenario, gender marking on adjectives and demonstrative would also follow from the emergence of number marking on demonstratives. When taking into account the other scenario, that is the evolutionary dynamics that lead to the co-occurrence of both agreement features (Figure 4b), gender agreement on demonstratives would seem to precede the emergence of number marking on both targets types. In addition, beyond strictly cross-target patterns, we also detect a strong tendency for gender agreement on adjectives to precede number agreement on adjectives. All in all, this suggests that, under both scenarios (the

emergence of agreement from scratch, and the evolutionary dynamics leading to the co-occurrence of both agreement features), demonstratives play a crucial role in fostering agreement marking within the noun phrase. In addition, when considering how languages arrive at the co-occurrence of both gender and number agreement, our findings suggest that the presence of gender agreement tends to call for more agreement to emerge.

3.2.3 Family-specific analyses

Finally, we consider the transition rate differences within specific language families for which phylogenies are available, i.e., Bantu, Dravidian, and Indo-European. Interestingly, we find no evidence for the tendency of gaining gender and number agreement within the same and across different targets in any family. This holds for rate differences in the analysis with both pair features absent as a starting point and with both pairs features present as the arrival point.

To summarize, we observe different tendencies depending on the perspective of gaining agreement. The global analysis suggests that languages with no marking tend to first develop number and then gender, whereas transitioning towards having both types of agreement typically happens when gender is already present and number agreement follows. When considering existing phylogenies from three large language families, we do not identify any significant trend between gender and number. This result may suggest that, for these specific language families, the evolution of gender and number agreement in the noun phrase may not be the most relevant domain of analysis when testing the implicational relationship posited by Greenberg. One could speculate that other agreement domains, such as for instance verbal or pronominal agreement, may be at stake in these languages. However, testing this hypothesis goes beyond the scope of the present paper.

4 Discussion

The goal of this study was twofold. Firstly, we aimed to test whether the implicational relation between gender and number posited by Greenberg (1963) holds when considering the morphosyntactic encoding of gender and number distinctions in the domain of noun-phrase internal agreement. This was achieved by observing the evolutionary dynamics of agreement, i.e. transition rates between types of agreement marking estimated on established phylogenies. Secondly, we set out to explore which types of agreement targets, between demonstratives, adjectives, and, at least in part, also articles, develop agreement first, and under which conditions. We conducted our quantitative analyses on a global sample first, to later delve into the

evolutionary dynamics of noun-phrase internal gender and number agreement in three distinct language families: Bantu, Dravidian, and Indo-European.

We distinguish two scenarios in the order of gaining number and gender agreement. In the first scenario, our findings support Greenberg's idea that gender marking hinges on number marking on the global level. Notably, we show that, in languages that lack either type of noun-phrase internal agreement, number agreement would tend to emerge before gender agreement. This first scenario could suggest that complex morphosyntactic phenomena, such as agreement marking, tend to emerge first in connection with cross-linguistically frequent and conceptually salient grammatical distinctions, as is the case for number. These findings also lend quantitative support to earlier claims made by Nichols (1992) and Wälchli et al. (2019), who suggested that for gender agreement to kick off, some form of agreement marking, usually in the domain of number and person, should already be present in the language.

The second scenario hinges on the evolutionary dynamics that lead to agreement marking of both categories. In such cases, the results of our global analyses show that, in the noun phrase domain, the development of gender agreement tends to precede that of number agreement. This second scenario, whereby the co-existence of the two types of agreement follows the presence of gender agreement, reflects the well-known fact that agreement is the most intrinsic property of gender systems. This is in contrast with the fact that, in many languages, nominal number marking is often restricted to nouns and does not involve any form of agreement. It also resonates with earlier findings by Di Garbo (2014), who observed that in the languages of Africa, pervasive agreement marking always involves both gender and number, but that, when a language lacks gender, number agreement is actually rare, notwithstanding the presence of number marking on nouns (see Section 1).

Far from being in contradiction with each other, the two scenarios of change outlined above contribute to build up a broader picture of the evolutionary dynamics of gender and number systems. On the one hand, we explore how a language starts developing noun-phrase internal agreement from scratch and find that number rather than gender is more likely to be developed first. On the other hand, we investigate how agreement in the noun phrase becomes more entrenched through the co-occurrence of both gender and number marking. In this latter case, the order in which both agreement types are gained is gender first, number second. Both of these findings lead us to the conclusion that number and gender agreement differ drastically in the degree of independence from each other. While number agreement is likely to develop in targets with and without gender agreement, gender depends on number in two distinct ways: 1) gender agreement is less likely to arise on targets without number agreement and 2) when it does happen, targets with gender agreement tend to develop number agreement.

In terms of the second goal of the study, exploring evolutionary trajectories of gender and number marking by type of target, the global analyses suggest that in the smaller sample of languages with articles, number agreement spreads from adjectives to demonstratives. In other words, adjectives develop number agreement before demonstratives do. This pattern is not found in the larger sample, where number agreement in demonstratives may instead follow or precede gender agreement in adjectives, depending on the evolutionary trajectory, that is, from the absence of both agreement features to the emergence of one or from the occurrence of gender agreement only to the presence of both. Both of these results align with recent studies on word order universals in the noun phrase. Based on a combination of experimental, typological, and corpus-based evidence, Culbertson et al. (2020) show that adjectives tend to be linearly closer to nouns because they are semantically tightly connected to nouns and their referents. The patterns emerging from the smaller sample of our global analyses could be interpreted as suggesting that, in line with the fact that adjectives are semantically tied to the nouns they modify and linearly closer to them, they would also be the first type of modifier to develop agreement. In this sense, agreement marking on adjectives would somewhat enhance the syntactic and semantic cohesion between nouns and their modifiers. The results from the larger sample could also be explained in terms of Culbertson et al. (2020). Given that demonstratives are less tightly associated with nouns in terms of mutual information, agreement marking on demonstratives may be seen as a means of overtly signaling this association with nouns. In other words, grammatical marking of syntactic cohesion within the noun phrase would tend to appear first on those modifiers that are less inherently associated with nouns, and as a way of overtly signaling this association in a given syntactic context. It seems to us that both interpretations would be directly testable through a behavioral experiment, the design of which goes beyond the scope of this paper. Finally, our results on the larger sample, which indicate the tendency of gender agreement on demonstratives preceding the emergence of number agreement on adjectives, are in line with the generalization proposed by Moravcsik (1997: 317), whereby the presence of adjective-noun agreement would imply the presence of demonstrative-noun agreement.

One limitation of our study is that the distinction between articles and demonstratives might not always be pronounced, and these categories might have identical or similar forms (Dryer 2013). The reason for the similarity of forms in some languages is that demonstratives represent the most common source construction for the grammaticalization of definite articles (Greenberg 1978). Further studies should explore in detail how exactly the forms of these targets might affect the evolution of agreement.

5 Concluding remarks

To our knowledge, the present study is the first of its kind to model the evolutionary trajectories of gender and number agreement on multiple types of agreement targets, based on a large dataset of languages of the world and with the support of cutting edge statistical methodologies. The global tendencies revealed by this study deserve to be studied in greater detail, by, for instance, expanding the domain of analysis beyond the noun phrase, that is, to the domain of verbal and pronominal agreement; or through behavioral experiments directly testing the explanations that we suggested in order to account for the observed distributions. Broadening the focus to other targets of agreement could shed light on family-specific trends, which were not identified in the scope of this paper. Given the current stage of documentation of verbal and pronominal agreement in large cross-linguistic databases, taking this step would require extensive data collection, which could be more readily achieved by focusing on individual language families first.

All in all, the study contributes to showcasing the benefits of using large-scale typological databases to test the validity of well-established typological generalizations for which empirical support was hitherto based only on small language samples.

Supplementary Materials

The code and the data used for the analysis can be downloaded at the following link: https://doi.org/10.17605/OSF.IO/WHRBC.

Acknowledgments: We wish to thank the editorial board of *Linguistic Typology* as well as three anonymous reviewers for their meticulous and constructive feedback. The usual disclaimers apply.
Research funding: O.S. received funding from the SALTO exchange programme between the CNRS and the MPG to support a 4-month research stay at the Ecological Anthropology lab (UMR 7206), Muséum National d'Histoire Naturelle (MNHN), Paris, France, during which the work contributing to this paper was conducted. M.A.-T. is grateful for the support of the French National Research Agency, grant EVOGRAM: The role of linguistic and non-linguistic factors in the evolution of nominal classification systems, ANR-20-CE27-0021. FDG is grateful to Aix-Marseille Université Recruitement Fund (MCVER80D), which sponsored the research collaboration culminating in this paper.

References

Barrett, Malcolm. 2021. ggokabeito: 'okabe-ito' scales for 'ggplot2' and 'ggraph', R package version 0.1.0. https://CRAN.R-project.org/package=ggokabeito.

Bouckaert, Remco R., Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, D. Russell Gray, Marc A. Suchard & Quentin D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097). 957–960.

Bouckaert, Remco R., David Redding, Oliver Sheehan, Thanos Kyritsis, Russell D. Gray, Kate E. Jones & Quentin D. Atkinson. 2022. Global language diversification is linked to socio-ecology and threat status. *preprint* https://osf.io/preprints/socarxiv/f8tr6/.

Burnham, Kenneth P. & David R. Anderson (eds.). 2004. *Model selection and multimodel inference*. New York, NY: Springer New York.

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

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

Corbett, Greville G. 2013. Number of genders. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/30.

Csardi, Gabor & Tamas Nepusz. 2006. The igraph software package for complex network research. *InterJournal Complex Systems* 1695. https://igraph.org.

Culbertson, Jennifer, Marieke Schouwstra & Simon Kirby. 2020. From the world to word order: Deriving biases in noun phrase order from statistical properties of the world. *Language* 96(3). 696–717.

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

Dryer, Matthew S. 2013. Definite articles (v2020.3). In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Gowri-Shankar, Vivek & Magnus Rattray. 2007. A reversible jump method for Bayesian phylogenetic inference with a nonhomogeneous substitution model. *Molecular Biology and Evolution* 24(6). 1286–1299.

Green, P. J. 1995. Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82. 711–732.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language: Report of a conference held at Dobbs Ferry, New York, April 13–15, 1961*, 73–113. Cambridge, Massachusetts: MIT Press.

Greenberg, Joseph H. 1978. *Universals of human language*. Stanford: Stanford Univ. Press.

Grollemund, Rebecca, Branford 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* 112(43). 13296–13301.

Hammarström, Harald, Robert Forkel & Martin Haspelmath & Sebastian Bank. 2021. Glottolog 4.5. http://glottolog.org.

Haspelmath, Martin. 2013. Occurrence of nominal plurality. In Matthew Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library Chapter 34.

Kassambara, Alboukadel. 2023. ggpubr: 'ggplot2' based publication ready plots. https://CRAN.R-project.org/package=ggpubr.Rpackageversion0.6.0.

Kay, Matthew. 2023. ggdist: Visualizations of distributions and uncertainty. *R package version 3.2.1*. https://mjskay.github.io/ggdist/.

Kolipakam, Vishnupriya, Fiona M. Jordan, Michael Dunn, Simon J. Greenhill, Remco R. Bouckaert, Russell D. Gray & Annemarie Verkerk. 2018. A Bayesian phylogenetic study of the Dravidian language family. *Royal Society Open Science* 5(3). 1–17.

Mace, Ruth & Clare J. Holden. 2005. A phylogenetic approach to cultural evolution. *Trends in Ecology & Evolution* 20(3). 116–121.

Macklin-Cordes, Jayden L. & Erich R. Round. 2022. Challenges of sampling and how phylogenetic comparative methods help: With a case study of the Pama-Nyungan laminal contrast. *Linguistic Typology* 26(3). 533–572.

Moravcsik, Edith. 1997. Parts and wholes in the Hungarian noun phrase-a typological study. In B. Palek (ed.), *Proceedings of LP '96, typology-prototypes, item orderings and universals*, 307–324. Prague: Charles University Press.

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

Pagel, Mark. 1994. Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 255(1342). 37–45.

Pagel, Mark, Andrew Meade & Daniel Barker. 2004. Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* 53(5). 673–684.

Paradis, Emmanuel & Klaus Schliep. 2018. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35. 526–528.

Paul, Waltraud. 2010. Adjectives in Mandarin Chinese: The rehabilitation of a much ostracized category. In Patricia Cabredo Hofherr & Ora Matushansky (eds.), *Linguistik Aktuell/Linguistics today*, vol. 153, 115–152. Amsterdam: John Benjamins.

Pedersen, Thomas Lin. 2022. ggraph: An implementation of grammar of graphics for graphs and networks. *R package version 2.1.0*. https://CRAN.R-project.org/package=ggraph.

R Core Team. 2023. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Raftery, A. 1996. Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika* 83(2). 251–266.

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

Schliep, K. P. 2011. Phangorn: Phylogenetic analysis in R. *Bioinformatics* 27(4). 592–593.

Schloerke, Barret, Di Cook, Joseph Larmarange, Francois Briatte, Moritz Marbach, Edwin Thoen, Amos Elberg & Jason Crowley. 2020. GGally: Extension to 'ggplot2'. *R package version 2.0.0*. https://CRAN.R-project.org/package=GGally.

Skirgård, Hedvig, Hannah J. Haynie, Damián E. Blasi, Harald Hammarström, Jeremy Collins, Jay J. Latarche, Jakob Lesage, Tobias Weber, Alena Witzlack-Makarevich, Sam Passmore, Angela Chira, Luke Maurits, Russell Dinnage, Michael Dunn, Ger Reesink, Ruth Singer, Claire Bowern, Patience Epps, Jane Hill, Outi Vesakoski, Martine Robbeets, Noor Karolin Abbas, Daniel Auer, Nancy A. Bakker, Giulia Barbos, Robert D. Borges, Swintha Danielsen, Luise Dorenbusch, Ella Dorn, John Elliott, Giada Falcone, Jana Fischer, Yustinus Ghanggo Ate, Hannah Gibson, Hans-Philipp Göbel, Jemima A. Goodall, Victoria Gruner, Andrew Harvey, Rebekah Hayes, Leonard Heer, Roberto E. Herrera Miranda, Nataliia Hübler, Biu Huntington-Rainey, Jessica K. Ivani, Marilen Johns, Erika Just, Eri Kashima, Carolina Kipf, Janina V. Klingenberg, Nikita König, Aikaterina Koti, Richard G. A. Kowalik, Olga Krasnoukhova, Nora L. M. Lindvall, Mandy Lorenzen, Hannah Lutzenberger, Tânia R. A. Martins, Celia Mata German, Suzanne van der Meer, Jaime Montoya Samamé, Michael Müller, Saliha Muradoglu, Kelsey Neely, Johanna Nickel, Miina Norvik, Cheryl Akinyi Oluoch, Jesse Peacock, India O. C. Pearey, Naomi Peck, Stephanie Petit, Sören Pieper, Mariana Poblete, Daniel Prestipino,

Linda Raabe, Amna Raja, Janis Reimringer, Sydney C. Rey, Julia Rizaew, Eloisa Ruppert, Kim K. Salmon, Jill Sammet, Rhiannon Schembri, Lars Schlabbach, Frederick W. P. Schmidt, Amalia Skilton, Wikaliler Daniel Smith, Hilário de Sousa, Kristin Sverredal, Daniel Valle, Javier Vera, Judith Voß, Tim Witte, Henry Wu, Stephanie Yam, Jingting Ye, Maisie Yong, Tessa Yuditha, Roberto Zariquiey, Robert Forkel, Nicholas Evans, Stephen C. Levinson, Martin Haspelmath, Simon J. Greenhill, Quentin D. Atkinson & Russell D. Gray. 2023. Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances* 9(16). eadg6175.

Soetaert, Karline. 2020. *Diagram: Functions for visualising simple graphs (networks), plotting flow diagrams*. https://CRAN.R-project.org/package=diagram. Rpackageversion1.6.5.

van der Bijl, Wouter. 2018. phylopath: Easy phylogenetic path analysis in r. *PeerJ* 6. e4718.

Verkerk, Annemarie, Olena Shcherbakova, Hannah J. Haynie, Hedvig Skirgård, Quentin D. Atkinson, Simon J. Greenhill & R. D. Gray. Forthcoming. Global support for some linguistic universals points to the importance of diachrony. (Accepted by *Nature Human Behaviour*).

Wälchli, Bernhard, Bruno Olsson & Francesca Di Garbo. 2019. *Grammatical gender and linguistic complexity, Volume 1: General issues and specific studies*. Language Science Press.

Wickham, Hadley. 2017. tidyverse: Easily install and load the Tidyverse. *R package version 1.2.1*. https://CRAN.R-project.org/package=tidyverse.

Xie, Wangang, Paul O. Lewis, Yu Fan, Lynn Kuo & Ming-Hui Chen. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology* 60(2). 150–160.