

# Evolutionary pathways of complexity in gender systems

Olena Shcherbakova<sup>1,\*</sup>,  and Marc Allassonnière-Tang<sup>2</sup>, 

<sup>1</sup>Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Deutsche Platz 6, Leipzig 04103, Germany

<sup>2</sup>EA UMR 7206, National Museum of Natural History, 17 place du Trocadéro, Paris 75016, France

\*Corresponding author: Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Deutsche Platz 6, Leipzig 04103, Germany. E-mail: [olena\\_shcherbakova@eva.mpg.de](mailto:olena_shcherbakova@eva.mpg.de)

Associate Editor: Dan Dediu

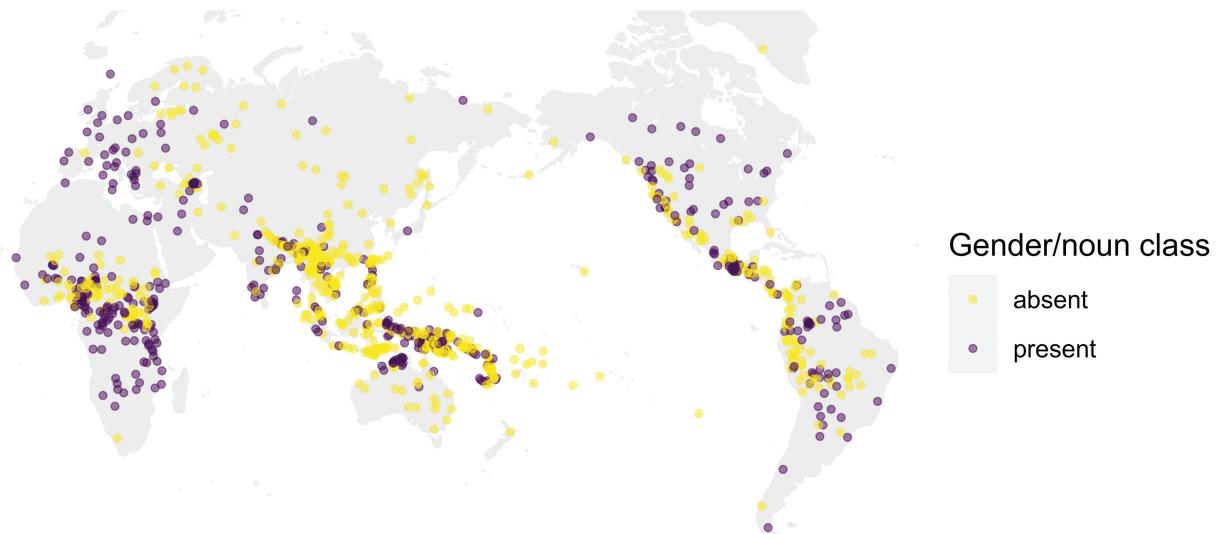
Humans categorize the experience they encounter in various ways, which is mirrored, for instance, in grammatical gender systems of languages. In such systems, nouns are grouped based on whether they refer to masculine/feminine beings, (non-)humans, (in)animate entities, or objects with specific shapes. Languages differ greatly in how many gender assignment rules are incorporated in gender systems and how many word classes carry gender marking (gender agreement patterns). It has been suggested that these two dimensions are positively associated as numerous assignment rules are better sustained by numerous agreement patterns. We test this claim by analyzing the correlated evolution (*Continuous method in BayesTraits*) and making the causal inferences about the relationships (phylogenetic path analysis) between these 2 dimensions in 482 languages from the global Grambank database. By applying these methods to linguistic data matched to phylogenetic trees (a world tree and individual families), we evaluate whether various types of gender assignment rules (semantic, phonological, and unpredictable) are causally linked to more gender agreement patterns on the global level and in individual language families. Our results on the world language tree suggest that semantic rules are weakly positively correlated with gender agreement and that the development of agreement patterns is facilitated by different rules in individual families. For example, in Indo-European languages, more agreement patterns are caused by the presence of phonological and unpredictable rules, while in Bantu languages, the driving force of agreement patterns is the variety of semantic rules. Our study shows that the relationships between agreement and rules are family-specific and yields support to the idea that more distinct rules and/or rule types might be more robust in languages with more pervasive gender agreement.

**Keywords:** grammatical gender; linguistic complexity; phylogenetics.

## 1. Introduction

One of the core tasks pursued by language sciences is to explain the remarkable linguistic diversity observed within more than 7,000 languages across the globe. Around a third of the world's languages (Allassonnière-Tang et al. 2021) have grammatical gender systems, which reflects the propensity of humans to categorize the world (Kemmerer 2014; Aikhenvald 2016; Kemmerer 2017a,b, 2019). These languages vary vastly in how they classify nouns into grammatical categories (Di Garbo and Miestamo 2019; Sinnemäki 2019). One of the most common types of nominal classification systems is gender/noun class systems that obligatorily group all nouns into distinct categories (Allan 1977;

Dixon 1986; Singer 2016). These systems differ greatly in the criteria that determine which nouns are grouped to which category and how deeply entrenched these systems are in the grammars. For instance, French and Italian nouns are either masculine or feminine. Such languages are typically said to have gender systems. Other languages such as Swahili have noun class systems that group nouns based on additional semantic criteria: animate/inanimate, human/non-human, animal/non-animal, plant/non-plant, as well as distinct shapes (e.g. long or round objects) or tools (Contin-Morava 2000). Areal and phylogenetic effects are also observed. As an example, most languages spoken in Europe have gender systems, while languages in Africa



**Figure 1.** The global distribution of gender (gender/noun class) systems based on the Grambank database. The points represent 1,151 languages included in the database.

tend to have either gender or noun class systems (Allasonnière-Tang et al. 2021). Following the terminology of the WALS (World Atlas of Language Structures) (Corbett 1991; Dryer and Haspelmath 2013), we use the term “gender” to refer to both gender and noun class systems (Fig. 1).

Gender systems are interesting because they hint toward the structure of the human cognitive system and the development of linguistic complexity. Categories found in such systems are not arbitrary. For example, categories based on the following distinction are frequently found in grammatical gender systems: **animate/inanimate**, **human/non-human**, **animal/non-animal**, **male/female**. Some shapes, such as long and round, are also commonly identified in gender systems. These tendencies match with the neuroscience premise that those categories are more salient in the human cognitive system and, therefore, more likely to be mirrored in human communicative systems (Kemmerer 2014, 2017a,b, 2019). The categories found in gender systems are also influenced by cognitive and cultural biases (Aikhenvald 2016; Kemmerer 2019). For example, the shape features of “long” and “round” are expected to be more common since they are salient shapes within the human cognitive system. These influences also reflect the observation that how nouns are affiliated to gender categories is also far from being arbitrary (Veeman et al. 2020; Allasonnière-Tang et al. 2021; Basirat et al. 2021).

Apart from vastly different semantic rules underlying gender assignment in the world’s languages, some

languages may additionally employ formal assignment rules (Dixon 1986). In such cases, the phonological or morphological form of the nouns rather than their meaning determines their gender. The famous mismatch between meaning and form of the German noun for “girl” (*das Mädchen*) assigned to neuter rather than feminine gender demonstrates the prevalence of the formal rule: the “*chen*” suffix triggers neuter gender on all nouns regardless of their meaning. This is uncommon since semantic rules typically prove to be decisive in gender assignment when semantic and formal rules are in conflict (Corbett 1991). For instance, in the Qafar language, even though the noun for “slender-waisted female” does not fit the phonological form of the feminine nouns (ending with an accented vowel), it still belongs to the feminine nouns (Corbett 2013).

Another distinctive property of gender systems is that they imply at least one agreement target: articles, demonstratives, verbs, adjectives, and pronouns may agree with nouns in gender and carry respective gender markers (Corbett 1991). Gender systems vary greatly in the extent to which gender is manifested on various agreement targets. For instance, in French, gender assignment is based on only one distinction (masculine and feminine), but its gender pervasiveness (the number of agreement patterns) (Liljegren 2019) is high. The articles, adjectives, participles, and demonstratives associated with a feminine noun are marked as feminine, c.f., *une grande lettre est écrite* (one.FEM big.FEM letter(FEM) is written.FEM) “a long letter has been written.” By contrast, the gender categories masculine

and feminine are not deeply entrenched in Assamese (Indo-Aryan), where only a few adjectives agree with nouns in gender (Bora 2004).

These two dimensions of variation in gender systems—number and/or type of gender assignment rules and number of agreement patterns—are also the criteria used to assess the complexity of these systems (along with the number of categories or values, e.g. German has three categories—feminine, masculine, and neuter) (Audring 2016; Di Garbo and Miestamo 2019). These dimensions have been claimed to interact with each other: “any change in the number of gender values or the number and nature of gender assignment rules must ultimately hinge on variation and change in the domain of agreement patterns” (Di Garbo and Miestamo 2019). Here, we explore whether two of these dimensions are interdependent, and we attempt to identify the constraints on the diversity of gender pervasiveness. Specifically, we test whether the number of assignment rules is positively or negatively associated with the number of agreement patterns. Additionally, we test the opposite scenario: whether the number of agreement patterns is influenced by the number of assignment rules.

## 2. Gender systems and linguistic complexity

Gender systems are treated as grammatically complex features primarily due to the obligatory grammatical marking associated with these systems (McWhorter et al. 2007). However, in some languages, the presence of gender systems will add to grammatical complexity when it is defined through other criteria, such as inflectional morphology, irregularity, or disruptions in meaning-form correspondences (Miestamo 2008; Lupyan and Dale 2010; Trudgill 2011). Since the presence of a gender system already contributes to the complexity of grammar, one could expect complexity in one dimension to be balanced out by simplicity in the other dimension, as suggested by the trade-off hypothesis. Nominal classification already has one prominent example of such a trade-off: the distribution of gender systems and numeral classifiers has been shown to be largely complementary (Sinnemäki 2019). In other words, when a language possesses one system of noun classification (gender), it is not economical to develop another system of the same type, such as numeral classifiers. Since a gender system is a complex phenomenon in itself (Dahl 2004; Trudgill 2011), languages could be less redundant if they compensated for extreme complexity in assignment rules by developing fewer agreement patterns. However, previous studies (Audring 2016; Di Garbo 2016) suggest that rules assignments

and agreement patterns should be positively correlated: simpler assignment rules co-occur with fewer agreement patterns, whereas more assignment rules imply more pervasive gender agreement. This positive association might be explained in the light of first language acquisition: children experience difficulties acquiring a rich gender system with numerous assignment rules if these are manifested in discourse via sparse agreement, for example, when the gender category is marked only on one target, such as an adjective (Audring 2016).

The available empirical evidence from synchronic studies of African languages supports the positive correlation between assignment rules complexity and the number of values (but not gender pervasiveness) (Di Garbo 2016). Many language families in Africa (e.g. Bantu) have diverse gender systems with a large number of gender values into which nouns are grouped based on humanness, animacy, plants, referential status, among others (Creissels and Pozdniakov 2015). Therefore, it is not clear whether the interdependence between different dimensions of gender complexity is prominent in languages beyond Africa. Another open question concerns different ways of measuring complexity of rules. In previous studies (e.g. Di Garbo 2016), the dimension of complexity is measured as a binary variable: simple assignment rules are purely semantic or formal, while complex rules rely on both semantic and formal criteria (but see Audring (2019) for a more nuanced proposal of distinguishing between semantic, phonological, and morphological assignment rules). However, due to a variety of semantic rules in the world’s languages, if more rules indeed imply more agreement patterns, the variety of semantic rules (rather than the combination of distinct rule types: semantic and formal vs purely semantic/formal) could potentially be a predictor of or be influenced by gender pervasiveness. Distinguishing between the number of semantic rules and the variety of distinct rule types (semantic and formal) is important because formal gender assignment rules never occur in isolation in the world’s languages (Corbett 1991). This means that most languages with gender systems group nouns based on either purely semantic rules or the combination of semantic and formal rules (Corbett 2013). In other words, if a language has formal rules, it is likely to also have semantic rules.

## 3. Materials and methods

Here, we assess if the potential evolutionary constraints on the diversity of gender systems are universal or family-specific (Dunn et al. 2011). To explore whether the complexity dimensions of gender systems are interdependent, we obtain the information on gender

rules and targets of agreement patterns from a large dataset of grammatical structures, Grambank (Skirgård et al. 2023), and model their evolution on the global tree (Jäger 2018) and typologically distinct language families with productive presence of gender systems: Austronesian (Gray et al. 2009), Bantu (Grollemund et al. 2015), Dravidian (Kolipakam et al. 2018), and Indo-European (Bouckaert et al. 2012).

### 3.1 Typological data

We use a large-scale database of typological features, Grambank (Skirgård et al. 2023), to obtain gender-related features for our analyses. First, we operationalize semantic rules as a continuous variable to test whether the variety of semantic rules alone can function as a predictor of the number of agreement patterns. Assessing assignment rules is especially challenging because “assignment rules may be hard to identify with certainty” and “[s]mall rules in particular are a source of disagreement among researchers” (Audring 2016). Due to this, the dimension of assignment rules is represented by four major semantic factors available in Grambank. These four binary features (GB051, GB052, GB053, GB054) listed in Table 1 are then aggregated to a score capturing whether the language has none (0), one (0.25), two (0.5), three (0.75), or all four (1) of these semantic factors in gender assignment. In many Indo-European languages, the commonly encountered rule is that the nouns are assigned to masculine or feminine gender based on their biological sex (GB051): for example, in French, nouns are either masculine or feminine. However, gender systems can incorporate other semantic rules. For instance, animacy can be a factor in gender assignment (GB053). In Grambank, this feature is coded as present also if the gender system captures a human versus non-human distinction. For instance, in Aneityu (Austronesian), one class of verbs takes different suffixes with animate (-*i*) and inanimate (-*ñ*) objects (Lynch 2000). Shape of the objects can also play a role in gender assignment (GB052). For instance, in Hulaulá (Afro-Asiatic), inanimate loanwords that denote long/thin entities, like feather or tail, are assigned the masculine gender (Khan 2009). Finally, in many languages, gender assignment distinguishes plants from other concepts (GB054). As an example, nouns denoting trees, plants as well as tree or plant parts fall under the third and fourth noun classes in Chuwabu (Atlantic-Congo). These same classes also host nouns referring to objects with a long, thin, and/or extended shape (Guérois 2015).

We also include two other rule types available in Grambank: phonological rules (GB192) and “unpredictable” (GB321) rules. Phonological rules are

**Table 1.** The Grambank features selected for this study. All features contributing to the aggregated scores of semantic rules complexity and agreement patterns complexity are assigned the same weight. The presence of one semantic rule weights 0.25 and the presence of one agreement pattern weights 0.2. The phonological and unpredictable rules features are binary and can be either present (1) or absent (0).

Feature ID	Features related to semantic rules
GB051	Is there a gender/noun class system where sex is a factor in class assignment?
GB052	Is there a gender/noun class system where shape is a factor in class assignment?
GB053	Is there a gender/noun class system where animacy is a factor in class assignment?
GB054	Is there a gender/noun class system where plant status is a factor in class assignment?
Features related to agreement patterns	
GB170	Can an adnominal property word agree with the noun in gender/noun class?
GB171	Can an adnominal demonstrative agree with the noun in gender/noun class?
GB172	Can an article agree with the noun in gender/noun class?
GB198	Can an adnominal numeral agree with the noun in gender/noun class?
GB030	Is there a gender distinction in independent 3rd person pronouns?
Binary features related to phonological and unpredictable rules	
GB192	Is there a gender system where a noun’s phonological properties are a factor in class assignment?
GB321	Is there a large class of nouns whose gender/noun class is not phonologically or semantically predictable?

treated as a binary variable since a more detailed information is often inaccessible: the descriptions of lesser known languages often lack formal assignment rules (phonological and morphological) (Audring 2016). The presence of phonological rules in language imply that gender assignment takes place based on the phonological form of the nouns, such as in Maltese (Afro-Asiatic), where most nouns ending in (-*a*) are feminine (Borg and Azzopardi-Alexander 1997). The so-called unpredictable rules refer to the presence of one or several gender/noun class categories whose members do not obviously share any semantic or phonological properties. For instance, three out of four genders are semantically defined in Mullukmulluk

(Northern Daly), whereas the fourth gender encompasses the rest of the nouns that do not meet the semantic criteria for inclusion into three other classes (Birk 1976).

The score of agreement patterns encompasses five features (GB170, GB171, GB172, GB198, GB030): agreement in gender of adnominal property words, demonstratives, articles, adnominal numerals, and the presence of gender distinctions in independent third person pronouns (four out of these five Grambank features are adnominal. With more data on other domains, in particular, agreement on verbs, future studies could overcome this limitation and test how alternative ways of measuring complexity of agreement patterns (e.g. assessing complexity by agreement domains rather than word class) respond to variation in assignment rules complexity.) In Ghomara Berber (Afro-Asiatic), adjectives of masculine and feminine nouns are distinguished based on their forms (GB170): adjectives of masculine nouns are unmarked whereas adjectives of feminine nouns take a specific suffix, which is also taken by adjectives of plural masculine and feminine nouns (el Hannouche 2008). Different demonstrative forms (GB171) are used with masculine, feminine, and inanimate nouns in the Australian language Ami (Western Daly). In Ami, demonstratives referring to inanimate nouns remain unmarked, and those that refer to masculine and feminine nouns get the suffixes *-na* and *-nga* respectively (Ford 1998). In Dutch, the definite singular article (GB172) of neuter nouns (*het*) is distinct from the singular/plural definite article of masculine and feminine nouns (*de*) (Donaldson 2008; Oosterhoff 2015). In Arbore (Afro-Asiatic), adnominal numerals agree with the nouns (GB198) in gender (masculine/feminine) (Hayward 1984). Finally, third person pronouns reflecting gender (GB030) are common in Indo-European languages such as in German (*er/sie/es*) and French (*lui/elle*). Third person pronouns can also have other gender distinctions. For example, in the pronominal paradigm of Aimele, a range of different prounoun forms distinguishing number and syntactic roles are used for animate reference, whereas the bound root *a-* is reserved for inanimate references regardless of number and syntactic roles (Aiton 2016).

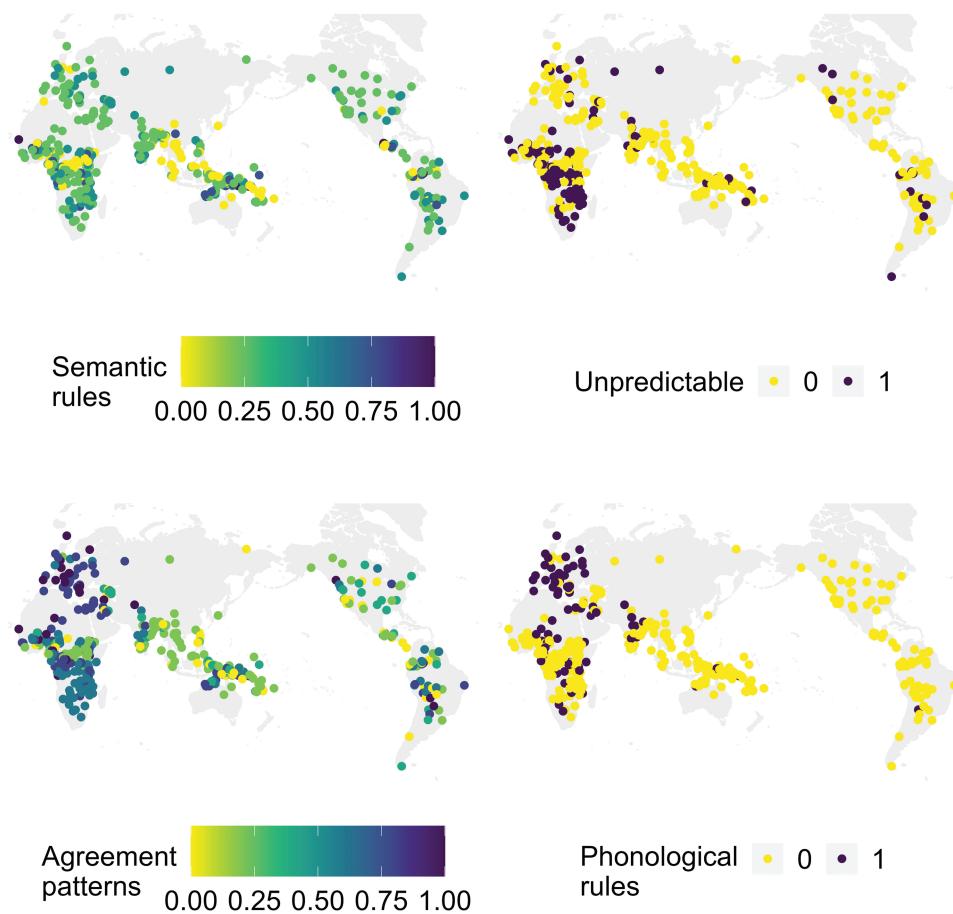
To summarize, the complexity of agreement patterns is calculated by aggregating the Grambank features dedicated to the agreement of demonstratives, adjectives, articles, numerals, and pronouns, with scores spanning the values of 0 (no marking at all) to 1 (all marked). The complexity of semantic rules is calculated in the same way, aggregating the presence of different assignment rules such as sex-based, shape-based, animacy-based, and plant-based.

Languages with missing data for at least one of these features were discarded. Two other rule types, phonological rules and unpredictable gender assignment of a large class of nouns, are coded as a binary feature indicating their presence/absence (the dependence/independence between these variables has not been fully investigated in existing studies, so that it remains unclear if some types of rules, such as animacy- and sex-based rules, might co-occur or be complementarily distributed. In the current study, we aggregate the variables assuming that they are independent and assign them the same weight. However, the exploration of the potential interactions between various rules (as well as between various agreement targets) would be a promising avenue for future research.) As shown in Fig. 2, hot spots of few semantic rules and few agreement patterns are situated in East and Central Africa as well as in Southeast Asia. Bantu and Indo-European languages exhibit highest gender pervasiveness (agreement scores). Phonological rules are mainly found in Europe, whereas unpredictable gender assignment typically occurs in Africa.

### 3.2 Phylogenies

The Indo-European tree accounts for 103 languages with a time depth of about 7,000 – 10,000 years, out of which 38 languages were used in the analysis. The Austronesian tree accounts for 400 languages with a time depth of about 5,000 years, out of which 27 languages matched with Grambank data. The Dravidian tree accounts for 20 languages with a time depth of about 4,500 years, and 12 of these were used in the analysis. The Bantu tree accounts for 425 languages with a time depth of about 5,000 years, out of which 67 languages matched with Grambank data. The languages from these language families cover over a third of the world's languages (Hammarström et al. 2022). We also use the global tree (Jäger 2018) that spans over 7,000 languages from 66 language families available in the ASJP database (Wichmann 2016) to infer potential global trends in the evolution of gender systems. We map 482 languages available in Grambank on the global tree. The distribution of semantic rules and agreement patterns are shown in Fig. 3.

Each tree represents a summary of the posterior tree samples for each of the language families and the world tree included in the analysis. Phylogenetic signal of continuous variables (semantic rules and agreement patterns) is estimated based on Pagel's lambda ( $\lambda$ ) (Pagel 1999) inferred with the help of *phylosig* function in *phytools* package (Revell 2012) (see Supplementary Table S2). The degree of support of the likelihood ratio (LR) test results is interpreted following Jeffreys



**Figure 2.** The global distribution of four variables describing two complexity dimensions of gender systems. Assignment rules are captured by semantic rules (continuous score), phonological rules (presence/absence), and unpredictable gender assignment of nouns with distinct semantic and phonological properties into the same class (presence/absence). Agreement patterns are coded as continuous score.

(1961). The phylogenetic signal of two binary traits (phonological and unpredictable rules) is measured as  $D$  values (Fritz and Purvis 2010) with the help of *phylo.d* function in *caper* package (Orme et al. 2013) (see Supplementary Table S3).

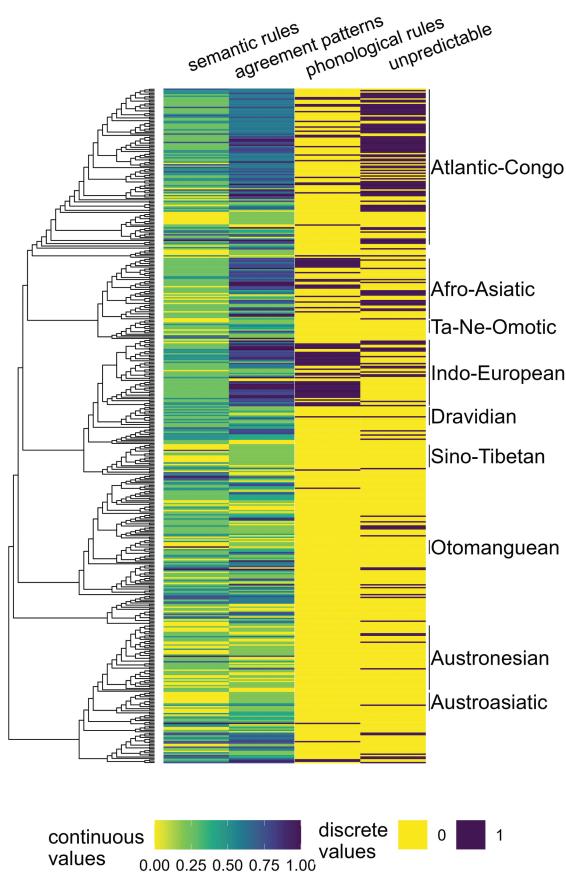
The typological data and the phylogenies are then combined for conducting phylogenetic analyses. First, we conduct phylogenetic path analysis (von Hardenberg and Gonzalez-Voyer 2013; van der Bijl 2018) to evaluate competing causal models where the number of agreement patterns influences or is driven by 1) the variety of semantic rules, 2) the presence of phonological rules, 3) the presence of “unpredictable” rules (the presence of a large or open set of nouns that are assigned to a single gender but do not share the same semantic or phonological properties), and 4) different combinations of these rules. This method enables

us to make inferences about the causal processes behind the expansion and reduction of agreement patterns and the changes in rule assignment. Second, we test for the presence of correlated evolution between two features—gender pervasiveness and the variety of semantic rules—using the *Continuous* method (analysis method: MCMC) implemented in *BayesTraits* (Pagel et al. 2004). Negative coevolutionary relationships will indicate that gender systems are constrained to be economical. A positive correlated evolution will suggest that there is evolutionary pressure for languages to accumulate redundancy in the gender systems.

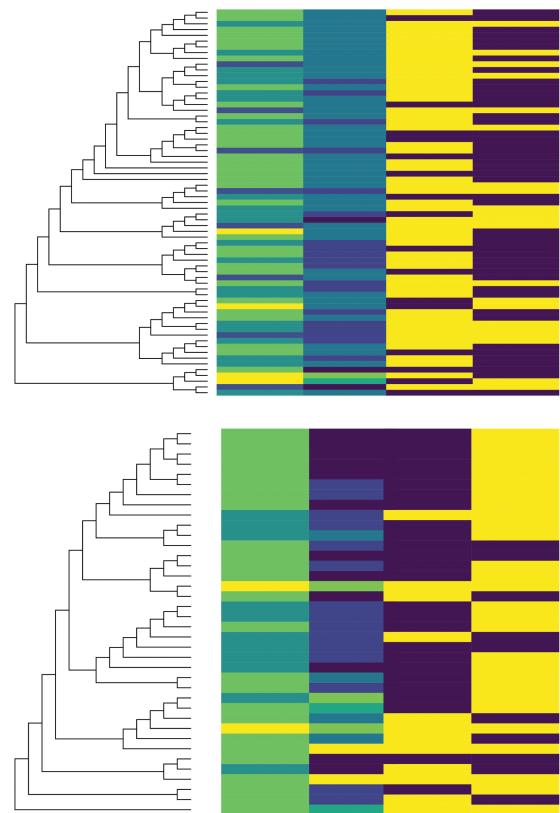
### 3.3 Causal relationship between agreement patterns and rules

To compare the effects of different assignment rules on agreement patterns and vice versa, we conduct

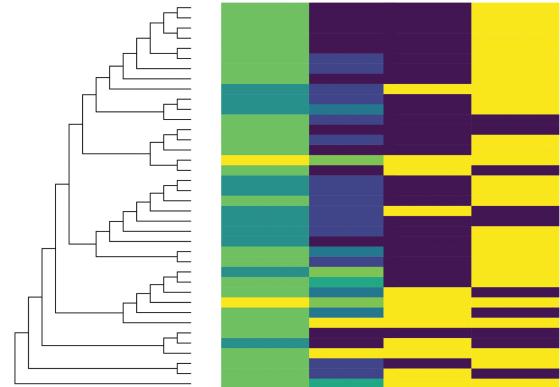
A



B



C



**Figure 3.** The values of the four variables mapped onto phylogenies. The depicted phylogenies were used for both phylogenetic path analysis and Continuous method: a) world tree, b) Bantu, and c) Indo-European. Unlike other large language families, Indo-European and Bantu show variation in the presence/absence of phonological and unpredictable rules. The presence of phylogenetic signal is attested for all four features on the global tree and for all features but unpredictable rules on Indo-European tree. The distribution of the features in Bantu does not seem to be phylogenetically constrained.

phylogenetic path analysis (von Hardenberg and Gonzalez-Voyer 2013) using the R package *phylopath* (van der Bijl 2018). This method applies d-sep test (Shipley 2009) and controls for phylogenetic non-independence by fitting phylogenetic generalized least-squares (PGLS) models (von Hardenberg and Gonzalez-Voyer 2013) (linear models and/or logistic regressions depending on the variables involved). Phylogenetic path analysis compares causal models in the form of DAGs (directed acyclic graphs) (Shipley 2000) based on the C-statistic information criterion corrected for small sample sizes (CICC) of each model. Model comparison is grounded in the idea that the minimum set of conditional independencies will be met only by the strongest model(s) (Shipley 2016). We develop a comprehensive set of competing causal models that

include different combinations of rule assignment predictors of agreement patterns with some models assuming that the predictors are also causally linked. The set also contains the causal models that predict that agreement patterns influence the number/presence of certain assignment rules (i.e. the reversed versions of the models where agreement patterns are predicted). In the future, the relationships between rules and agreement patterns could also be explored using exploratory path analysis. For instance, the BEPA (Brute-force exploratory path analysis) R package (<https://github.com/Joseph-Watts/BEPA>) builds upon the *phylopath* package van der Bijl (2018) and is available at the moment of writing in the beta version. The approach does not require the set of prespecified models and explores all potential combinations of

causal models, while allowing the exclusion of models containing specific causal paths. The limitation is that this approach is more time-intensive than testing a set of predefined competing models and so far has been tested for the maximum of six variables.

### 3.4 Coevolution of agreement patterns and rules

We use the Bayesian phylogenetic approach to evaluate the coevolution between semantic rules and agreement patterns. We focus on these two continuous variables because it allows us to 1) test whether changes in gender pervasiveness depend on the changes in the number of assignment rules on further language families with gender systems (Dravidian ([Kolipakam et al. 2018](#)) and Austronesian ([Gray et al. 2009](#))) where phonological and unpredictable rules are predominantly absent and 2) cross-check our results of phylogenetic path analysis on Indo-European and Bantu phylogenies with another method that also controls for phylogenetic non-independence. We use the *Continuous* method implemented in the *BayesTraits* software ([Pagel et al. 2004](#)) to assess how the two variables interact with each other diachronically.

We fit two models (see [Supplementary Tables S2](#) and [S3](#) for an overview of the phylogenetic signal in the features). This allows us to compare the models of dependent and independent evolution. Under the first model, the covariance between two traits is estimated, whereas the second model assumes that two traits have evolved independently and their covariance is equal zero. To determine which of the model fits the data better, we calculated Bayes Factors ([Burnham and Anderson 2002](#)) from the marginal likelihoods of both models that we obtained using a stepping stone sampler ([Xie et al. 2011](#)) with 1,000 stones and 10,000,000 iterations per stone. The Bayes Factor is estimated in the following way:  $2 \times (\log \text{marginal likelihood of dependent model} - \log \text{marginal likelihood of independent model})$ . We interpret Bayes Factors above 2 as weak evidence, above 5 as strong, and above 10 as very strong evidence in support of the dependent model ([Raftery 1996](#)).

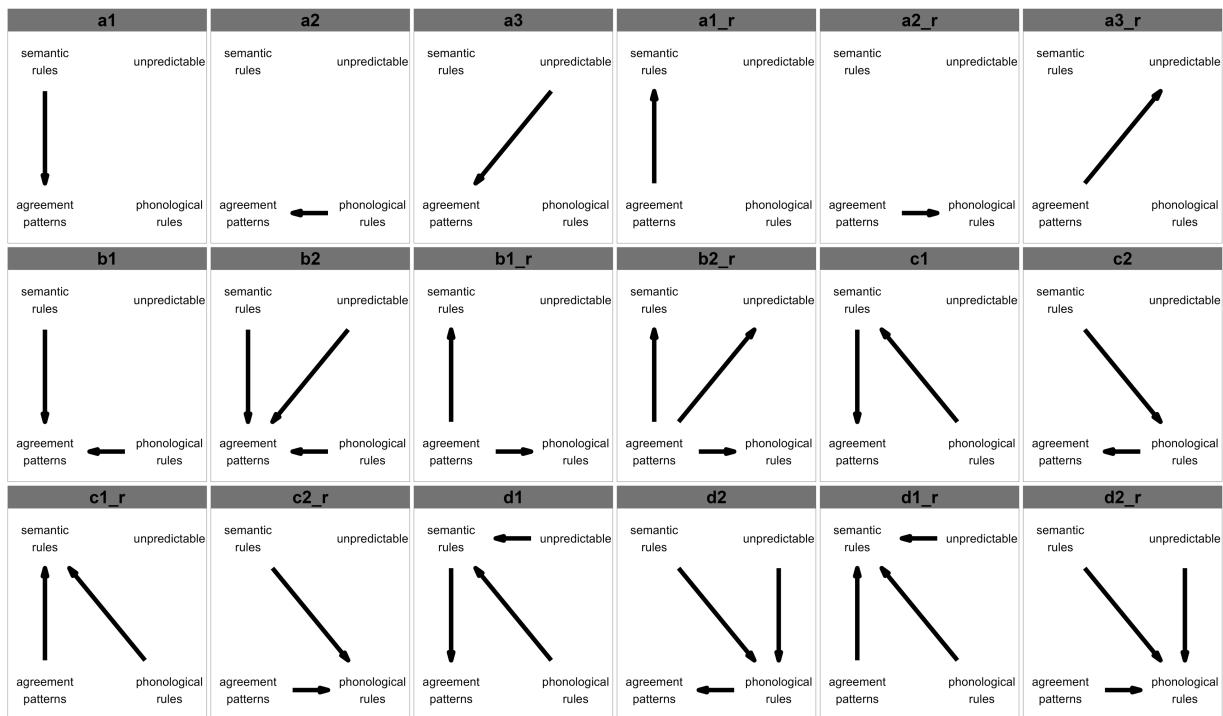
In order to avoid the possibility of running the analyses (both with the help of phylogenetic path analysis and the *Continuous* method) on the sample including languages without gender, we discard languages if they have neither semantic rules nor agreement patterns (i.e. if they score “0” on both dimensions). This also means that we eliminate the languages that might have gender systems based on other rules and mark gender on other targets than specified in Grambank. This reduces the size of our sample but allows for more robust results. Specifically, this prevents inflated positive correlation

due to many languages scoring 0 for both semantic rules and agreement patterns. We provide the results of Continuous analysis on the individual trees that contain languages with 0 scores for both dimensions (see [Supplementary Table S5](#)).

## 4. Results

Phylogenetic path analysis allows us to test whether agreement patterns predict or are sensitive to different assignment rules: 1) semantic, 2) formal (phonological), and 3) unpredictable. We run the analyses on the global sample as well as separately on Indo-European and Bantu languages. Unlike other families, where phonological and unpredictable rules are mainly absent, languages in these two language families vary in how the values of these rule features are distributed. We compile a comprehensive set of causal models belonging to four groups (see [Fig. 4](#)). Group (a) encompasses three simple models where agreement patterns depend on one out of three rule types. Group (b) explores the effects of different rule types on agreement patterns within the same models. Group (c) incorporates scenarios where agreement patterns arise due to more semantic rules/presence of phonological rules and different rule types are interdependent. This group of models accounts for the relationship that might arise from languages having either semantic or semantic and phonological assignment rules. Group (d) expands on the models in group (c) by adding the potential causal link between semantic/phonological rules and unpredictable rules. This allows us to detect potential patterns of loss of some semantic/phonological rule, which resulted in a synchronically observable large class of nouns with distinct semantic and phonological properties. Each causal model belonging to these four groups has a reverse version where agreement patterns are not longer outcome variables, but predictors of assignment rules.

Phylogenetic path analysis confirms that agreement patterns and gender assignment rules are positively correlated. Based on two top-ranking models on the global tree (b2 and its reverse version), we find no evidence for directionality of change between semantic rules and agreement rules on the global level: the effects of agreement patterns on semantic rules and semantic rules on agreement patterns are positive and weak. The causal relationships between agreement patterns and phonological rules similarly emerge as bidirectional, with agreement patterns facilitating the presence of unpredictable rules being more likely than the reverse. Agreement patterns have also been found to positively influence the emergence of phonological rules globally. Intriguingly, the effects of different rule types



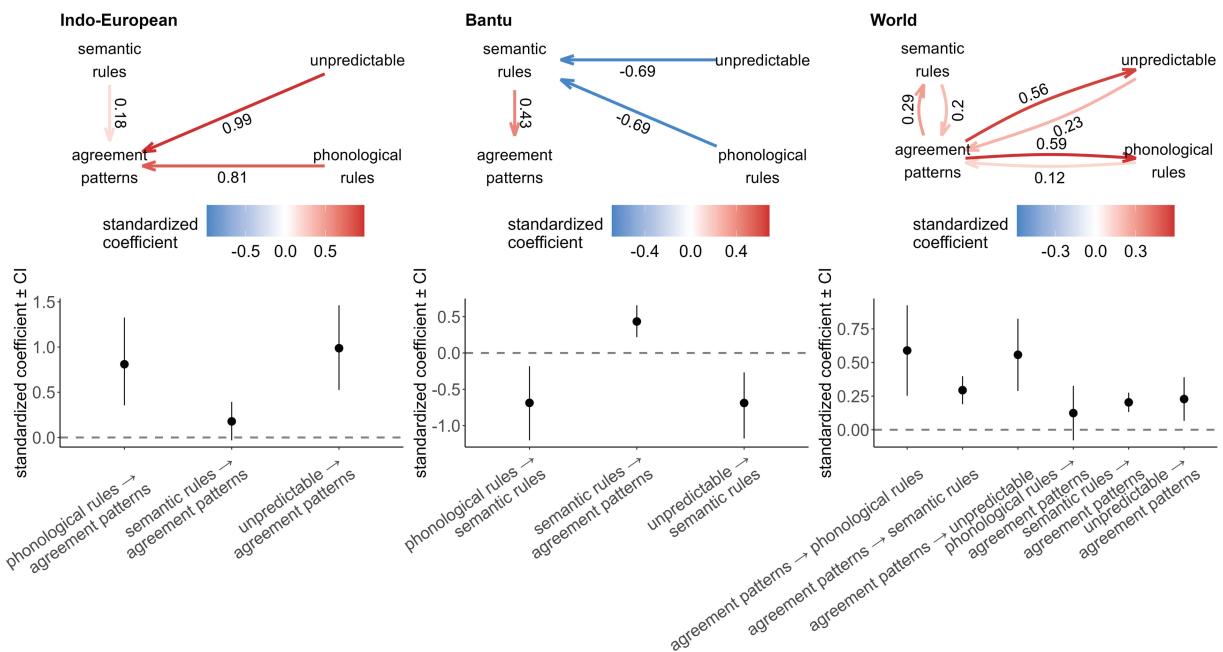
**Figure 4.** Eighteen competing causal models tested on Indo-European, Bantu, and global phylogenies within phylogenetic path analysis. The arrows represent the direction of causal paths.

vary across families (see Fig. 5). For the phylogenies of Indo-European and Bantu languages, the analysis establishes one strongest model, which fits the data considerably better than other models (see Supplementary Table S1). The difference between C-statistic information criterion corrected for small sample sizes (CICC) of the best models and the second-best models is  $> 7$ . Low ( $<0.05$ )  $P$  values for all but the strongest models allow to reject the rest of the models based on the available evidence.

In Bantu languages, the relationship between assignment rules and agreement patterns is best captured by the strongest model (d1, weight = 0.98) that implies the interrelationships between different rule types: the presence of phonological and unpredictable rules decreases the number of semantic rules, and numerous semantic rules lead to higher agreement patterns. The presence of phonological and unpredictable rules in Bantu languages do not seem to influence gender pervasiveness but makes abundant semantic rules unlikely. By contrast, the strongest causal model (b2) on the Indo-European (weight = 0.95) and global (weight = 0.95) phylogenies assumes the interaction between different rule types and agreement patterns and no interrelationships between rule types: languages

with more semantic rules and available phonological and unpredictable rules have more agreement patterns. However, the effects of some rule types on agreement patterns are not robust. For instance, the confidence intervals (CI) of semantic rules on agreement patterns on the Indo-European tree and phonological rules on agreement patterns on the global tree cross zero. This means that agreement patterns are triggered only by some rule types: in Indo-European languages these are caused by the presence of phonological and unpredictable rules, while the global tendency is that agreement patterns increase in languages with more semantic and present unpredictable rules.

In terms of correlated evolution, the results obtained with the *Continuous* method within *BayesTraits* (Pagel et al. 2004) yield weak (Bayes factor = 2.61 for Indo-European) and strong (Bayes factor = 11.56 for Bantu and Bayes Factor = 26 for the world tree) support to the dependent model of evolution between semantic rules and agreement patterns scores across the phylogenies (see Supplementary Table S4). The positive correlation between the traits is strongest on the Bantu tree ( $r = 0.4$ ), while on Indo-European ( $r = 0.26$ ) and global ( $r = 0.23$ ) phylogenies this relationship is weaker. Calculated Bayes factors yield no support in



**Figure 5.** The results of phylogenetic path analysis on three phylogenies. The standardized coefficients with CI not crossing 0 illustrate positive correlation between the number of rules/rule types and agreement patterns. One strong model explaining the constraints on gender pervasiveness was established in each phylogeny. The number of agreement patterns is driven by 1) the presence of phonological and unpredictable rules in Indo-European, 2) the number of semantic rules in Bantu (the score of which is, in turn, lower if a language already has phonological or unpredictable rules) and 3) bidirectionality in the global tree.

favor of the dependent models for Dravidian (Bayes factor = 0.21) and Austronesian (Bayes factor = 0.49) language families. Thus, while we find evidence for the coevolution between semantic rules and agreement patterns on the global level, such positive coevolutionary relationship is lineage-specific (Dunn et al. 2011) and evident only on some (Bantu and Indo-European) language families.

The results of two methods largely overlap and reveal positive correlation between semantic rules and agreement patterns on the Bantu and on the global phylogenies. On these samples, semantic rules remain positively correlated with agreement patterns after other rule types are included in the models. The only discrepancy concerns the findings on the Indo-European phylogeny: the *Continuous* method provides weak evidence for the weak positive correlation, whereas in phylogenetic path analysis, this predictor is not influential and agreement patterns are instead constrained by the presence/absence of phonological and unpredictable rules. This shows that not accounting for other rule types makes the weak positive correlation suggested by the *Continuous* method disappear.

## 5. Concluding discussion

Our global study goes beyond pairwise correlations that could potentially yield spurious results (Blasi and Roberts 2017) and examines the effects of multiple predictors of gender pervasiveness in 482 languages. Our analyses are based on the phylogenetic causal graph methods that account for genealogical relatedness between the languages in our sample (see Fig. 3) and reveal the causal processes behind the current feature distribution. The results generally show a positive correlation between the number of semantic rules and the number of agreement patterns. These results indicate that gender systems are not constrained to be economical, as we do not find a negative coevolutionary dependency between the assignment rules and agreement patterns. These results favor the explanation based on first-language acquisition and the claim that **gender systems with multiple rules and agreement patterns might be easier to acquire**. However, the positive correlation between semantic rules and agreement patterns is not equal across language families and the world tree. A moderate relationship is found for the Bantu languages. This observation

matches with existing studies (Di Garbo 2016). Nevertheless, we establish no relationship between these features in Indo-European, Austronesian, or Dravidian languages. Taken together, these findings show that the relationships between semantic rules and agreement patterns is lineage-specific. However, the analyses based on the world tree reveals a weak positive correlation. This can be explained by the fact that the world tree includes further language families that we have separately analyzed here. It could be that the correlation between semantic rules and agreement patterns is strong on other language families, but we did not include them due to the lack of specialist-established family trees. It might be that there is a bias toward combinations of more agreement patterns with more semantic rules (and fewer agreement patterns with semantic rules) in language families (or language family branches) that show substantial variation in the number of semantic rules, which is the case with Bantu languages, some of which have three out of four semantic rules based on our data. The development of other phylogenies is thus needed to further test these claims.

The phylogenetic path analysis results highlight the importance of accounting for multiple variables while controlling for the non-independence of languages in the sample. Our analyses show that the weak positive coevolutionary relationship between gender agreement and semantic assignment rules does not hold in Indo-European when other assignment rules are taken into account. Instead, the agreement patterns in this language family are driven to a great extent by the rules grounded in phonological properties and those that are phonologically and semantically unpredictable. This is in line with previous work showing that languages with formal and semantic assignment rules are more likely to have pervasive agreement as opposed to languages with purely semantic assignment rules (Di Garbo 2016). In that study, the rule assignment variable was binarized to reflect whether a language had a purely semantic/formal assignment rule or combined both semantic and formal assignment rules.

Similarly, using an explicit causal approach allows us to avoid the potential pitfalls of including more variables than necessary, which can, in some cases, undermine the analyses due to the “collider bias” (McElreath 2020). Our results suggest that semantic rules in the best-supported model for Bantu languages act as a collider, a variable influenced by multiple variables: phonological and unpredictable rules, both of which are negatively correlated with semantic rules. From this causal structure, we can infer that phonological and unpredictable rules are not correlated with each other. However, fitting a model where phonological rules serve as an outcome and unpredictable

and semantic rules as predictors would result in a negative but spurious correlation between phonological and unpredictable rules. This can leave an misleading impression that Bantu languages tend to have either phonological or unpredictable rules (a causal path where unpredictable rules influence phonological rules was also tested in our model set but not confirmed as supported), whereas the correct implication is that two variables are not causally connected but the presence of either is associated with fewer semantic rules.

Notably, we find no evidence for complexity trade-offs and all discovered substantial relationships between rules and agreement patterns are invariably positive. This indicates that gender systems are not constrained to be economical but are predominantly shaped by the needs of first-language speakers. When a language develops a gender system, especially one with numerous rules, it might be more robustly transmitted to future generations if the complexity of rules is comparable to the complexity of agreement patterns. Conversely, the process of first language acquisition of gender systems in languages with few rules and few agreement patterns, like Dutch, might take longer time because with sparse gender agreement, children encounter fewer opportunities to acquire the existing rules (Audring 2014). This is in line with the view that gender systems serve a range of lexical and discourse functions (Contini-Morava and KilarSKI 2013), such as error-checking (Dahl 2004) and reference-tracking (Corbett 1991), rather than being redundant and “unnecessary to human communication” (McWhorter 2001).

However, we do uncover two unexpected instances of trade-offs between the variety of semantic rules and the presence of other rule types in Bantu languages. Specifically, languages with phonological or unpredictable rules are more likely to have fewer semantic rules. These trade-offs appear to be restricted to Bantu languages and are not discovered on the global tree or in Indo-European languages. It might be that such relationships between different rule types develop in languages like Bantu which generally possess numerous semantic rules. Alternatively, in Bantu languages, the development of additional rule types might be more closely associated with the replacement of the available semantic rules.

Apart from the first-language acquisition hypothesis (Audring 2014) that posits the positive relationship between different complexity dimensions of gender systems, more agreement patterns might be found in languages with complex assignment rules for another reason. In some languages, distinct rules come to be associated with different agreement targets. For instance, some varieties of Pashayi (Indo-Aryan) have gender

systems with sex- and animacy-based assignment rules. While most agreement patterns found in these languages reflect the masculine-feminine differentiation, animacy-based distinctions are present exclusively in the verbal paradigm (Liljegren 2019). This way, additional rules might also imply additional agreement patterns.

Our results should be interpreted keeping the limitations of the grammatical data and methods in mind. We used typological information from Grambank (Skirgård et al. 2023) to obtain information on assignment rules and agreement patterns. However, gender-related features presented in this database are not exhaustive and do not cover all possible gender assignment rules and agreement patterns. For instance, we do not have separate features on human versus non-human semantic distinction or gender agreement of verbs. Further work should explore whether more detailed cross-linguistic studies corroborate the positive associations revealed here. Besides, we do not account for the effects of contact in our analyses, which could also shed light on the evolution of complexity in gender systems. This might be especially relevant for the Bantu languages where the distribution of gender features did not have strong phylogenetic signal. At the same time, when analyzing the properties of gender systems, controlling for phylogenetic autocorrelation might be sufficient given that gender systems features are among features resistant to borrowing (Greenhill et al. (2017); Nichols (2003); Allasonnière-Tang and Dunn (2020); Allasonnière-Tang et al. (2021); and Stoltz and Levkovych (2022)).

Finally, our results contribute to the existing studies claiming that positive complexity correlations can be diachronically well-motivated, as it is also found between syllable complexity and morphological synthesis (Easterday et al. 2021) and grammatical marking of nominal words and verbs in Sino-Tibetan languages (Shcherbakova et al. 2022). Our study suggests that the evolution of gender systems complexity is shaped by the drive for robustness rather than the economy constraints. Further work should investigate the influence of other potential predictors of gender agreement patterns in other language families and how the third complexity dimension (gender values) interacts with complexity of rules and agreement patterns on the global scale and in individual families.

## Acknowledgments

The authors are grateful to Andrew Meade for his advice on the methods. Marc Allasonnière-Tang is thankful for the support from the French National Research Agency (ANR-20-CE27-0021).

Special thanks to the editors and the reviewers who helped to improve the quality of the paper.

## Supplementary data

Supplementary data is available at *Journal of Language Evolution Journal* online.

## Conflict of interests

The authors declare no conflict of interests.

## Funding

This work was supported by the Department of Linguistic and Cultural Evolution (to O.S.) and the French National Research Agency (to M.A.-T., grant EVOGRAM: The role of linguistic and non-linguistic factors in the evolution of nominal classification systems, ANR-20-CE27-0021).

## Data availability

This article contains Supplementary Materials. Additionally, the code and data for reproducing this study can be found at <https://zenodo.org/doi/10.5281/zenodo.10559003>.

## References

- Aikhengvald, A. Y. (2016) *How Gender Shapes the World*. Oxford: OUP.
- Aiton, G. (2016) ‘A Grammar of Eibela: A Language of the Western Province, Papua New Guinea’, PhD thesis, James Cook University, Douglas, Queensland, Australia.
- Allan, K. (1977) ‘Classifiers’, *Language*, 53:285–311.
- Allasonnière-Tang, M., Brown, D., and Fedden, S. (2021) ‘Testing Semantic Dominance in Mian Gender: Three Machine Learning Models’, *Oceanic Linguistics*, 60:302–34.
- Allasonnière-Tang, M., and Dunn, M. (2020) ‘The Evolutionary Trends of Grammatical Gender in Indo-Aryan Languages’, *Language Dynamics and Change*, 11:211–40.
- Allasonnière-Tang, M. et al. (2021) ‘Expansion by Migration and Diffusion by Contact is a Source to the Global Diversity of Linguistic Nominal Categorization Systems’, *Humanities and Social Sciences Communications*, 8:331.
- Audring, J. (2014) ‘Gender as a Complex Feature’, *Language Sciences*, 43:5–17.
- Audring, J. (2016) ‘Gender’, in M. Aronoff (ed.) *Oxford Research Encyclopedia of Linguistics*. Oxford: OUP. <https://doi.org/10.1093/acrefore/9780199384655.013.43>.
- Audring, J. (2019) ‘Canonical, Complex, Complicated?’, in F. Di Garbo, B. Olsson, and B. Walchli (eds) *Grammatical Gender and Linguistic Complexity*, Vol. 1, pp. 15–52. Berlin, Germany: Language Science Press.
- Basirat, A., Allasonnière-Tang, M., and Berdicevskis, A. (2021) ‘An Empirical Study on the Contribution of Formal and

- Semantic Features to the Grammatical Gender of Nouns,’ *Linguistics Vanguard*, 7:20200048.
- Birk, D. B. W. (1976) ‘The MalakMalak language, Daly River (Western Arnhem Land)’, *Pacific Linguistics: Series B*, Vol. 45. Canberra: Australian National University. <http://sealang.net/archives/pl/pdf/PL-B45.pdf>.
- Blasi, D. E., and Roberts, S. G. (2017) ‘Beyond Binary Dependencies in Language Structure’, in N. J. Enfield (ed.) *Dependencies in Language: On the Causal Ontology of Linguistic Systems*, pp. 117–28. Berlin, Germany: Language Science Press.
- Bora, L. (2004) ‘Gender System in Assamese and Boro: A Contrastive Analysis’, in S. Burusphat (ed.), *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society*, pp. 635–43. Arizona State University.
- Borg, A., and Azzopardi-Alexander, M. (1997) *Maltese Descriptive Grammars Series*. London: Routledge.
- Bouckaert, R. et al. (2012) ‘Mapping the Origins and Expansion of the Indo-European Language Family’, *Science*, 337: 957–60.
- Burnham K. P., and Anderson D. R. (2002) *Model Selection and Multimodel Inference*. New York, NY: Springer-Verlag.
- Contini-Morava, E. (2000) ‘Noun Class as Number in Swahili’, in E. Contini-Morava, and Y. Tobin (eds) *Current Issues in Linguistic Theory*, Vol. 183, pp. 3–30. Amsterdam: John Benjamins Publishing Company.
- Contini-Morava, E., and Kilarski, M. (2013) ‘Functions of Nominal Classification’, *Language Sciences*, 40:263–99.
- Corbett G. G. (1991) *Gender*. Cambridge: CUP.
- Corbett G. G. (2013) ‘Systems of Gender Assignment’, in M. S. Dryer, and M. Haspelmath (eds) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Creissels, D., and Pozdniakov, K. I., eds (2015) ‘Les classes nominales dans les langues atlantiques’, *Grammatische Analysen afrikanischer Sprachen [Grammatical Analyses of African languages]*, Vol. 49. Köln: Rüdiger Köppe Verlag. OCLC: ocn944432140.
- Dahl, Ö. (2004) *The Growth and Maintenance of Linguistic Complexity*, Vol. 71. John Benjamins Publishing.
- Di Garbo, F. (2016) ‘Exploring Grammatical Complexity Crosslinguistically: The Case of Gender,’ *Linguistic Discovery*, 14:46–85.
- Di Garbo, F., and Miestamo, M. (2019) ‘The Evolving Complexity of Gender Agreement Systems’, in F. Di Garbo, B. Olsson, and B. Wälchli (eds) *Grammatical Gender and Linguistic Complexity*, Vol. 2, pp. 15–60. Language Science Press.
- Dixon, R. M. W. (1986) ‘Noun Class and Noun Classification’, in C. Craig (ed.) *Noun Classes and Categorization*, pp. 105–12. Amsterdam: John Benjamins.
- Donaldson, B. (2008) *Dutch: A Comprehensive Grammar*, 2nd edn. London: Routledge.
- Dryer, M. S., and Haspelmath, M. (2013) WALS Online. Leipzig, Germany: Max Planck Institute for Evolutionary Anthropology.
- Dunn, M. et al. (2011) ‘Evolved Structure of Language Shows Lineage-Specific Trends in Word-Order Universals’, *Nature*, 473:79–82.
- Easterday, S. et al. (2021) ‘Syllable Complexity and Morphological Synthesis: A Well-Motivated Positive Complexity Correlation Across subdomains’, *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.638659>
- el Hannouche, J. (2008) ‘Ghomara Berber: A Brief Grammatical Survey’, Master’s thesis, Rijksuniversiteit te Leiden.
- Ford, L. J. (1998) ‘A Description of the Emmi language of the Northern Territory of Australia’, PhD thesis, Australian National University, Canberra.
- Fritz, S. A., and Purvis, A. (2019) ‘Selectivity in Mammalian Extinction Risk and Threat Types: A New Measure of Phylogenetic Signal Strength in Binary Traits’, *Conservation Biology*, 24:1042–51.
- Gray, R. D., Drummond, A. J., and Greenhill, S. J. (2009) ‘Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement’, *Science*, 323:479–83.
- Greenhill S. J. et al. (2017) ‘Evolutionary Dynamics of Language Systems’, *Proceedings of the National Academy of Sciences*, 114:E8822–9.
- Grollemund, R. et al. (2015) ‘Bantu Expansion Shows that Habitat Alters the Route and Pace of Human Dispersals’, *Proceedings of the National Academy of Sciences*, 112:13296–301.
- Guéris, R. (2015) ‘A Grammar of Cuwabo (Mozambique, Bantu P34)’, PhD thesis, Université Lumière Lyon 2.
- Hammarström, H. et al. (2022) *Glottolog 4.6*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- von Hardenberg, A., and Gonzalez-Voyer, A. (2013) ‘Disentangling Evolutionary Cause-Effect Relationships With Phylogenetic Confirmatory Path Analysis’, *Evolution: International Journal of Organic Evolution*, 67: 378–87.
- Hayward, D. (1984) *The Arbore Language: A First Investigation Including a Vocabulary*. Kuschitische Sprachstudien 2. Cologne: Rüdiger Köppe.
- Jeffreys, H. (1961) *Theory of Probability*, 3rd edn. Oxford: OUP.
- Jäger, G. (2018) ‘Global-Scale Phylogenetic Linguistic Inference from lexical Resources’, *Scientific Data*, 5: 180189.
- Kemmerer, D. (2014) ‘Word Classes in the Brain: Implications of Linguistic Typology for Cognitive Neuroscience’, *Cortex*, 58:27–51.
- Kemmerer, D. (2017a) ‘Categories of Object Concepts Across Languages and Brains: The Relevance of Nominal Classification Systems to Cognitive Neuroscience’, *Language, Cognition and Neuroscience*, 32:401–24.
- Kemmerer, D. (2017b) ‘Some Issues Involving the Relevance of Nominal Classification Systems to Cognitive Neuroscience: Response to Commentators’, *Language, Cognition and Neuroscience*, 32:447–56.
- Kemmerer, D. (2019) *Concepts in the Brain: The View From Cross-linguistic Diversity*. Oxford: OUP.
- Khan, G. (2009) *The Jewish Neo-Aramaic Dialect of Sanandaj*. Piscataway: Gorgias.
- Kolipakam, V. et al. (2018) ‘A Bayesian Phylogenetic Study of the Dravidian Language Family’, *Royal Society Open Science*, 5:1–17.
- Liljegren, H. (2019) ‘Gender Typology and Gender (in) Stability in Hindu Kush Indo-Aryan Languages’, in F. Di Garbo,

- B. Olsson, and B. Wälchli (eds) *Grammatical Gender and Linguistic Complexity*, Vol. 1, pp. 279–328. Language Science Press.
- Lupyan, G., and Dale, R. (2010) ‘Language Structure Is Partly Determined by Social Structure’, *PloS One*, 5:e8559.
- Lynch, J. (2000) *A Grammar of Anejom: Pacific Linguistics*, Vol. 507. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- McElreath, R. (2020) *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. New York: Chapman and Hall/CRC.
- McWhorter, J. et al. (2007) *Language Interrupted: Signs of Non-native Acquisition in Standard Language Grammars*. New York, NY: OUP.
- McWhorter, J. H. (2001) ‘The Worlds Simplest Grammars Are Creole Grammars,’ *Linguistic Typology*, 5:125–66.
- Miestamo, M. (2008) ‘Grammatical Complexity in a Cross-Linguistic Perspective’, in M. Miestamo, K. Sinnemäki, and F. Karlsson (eds) *Language Complexity: Typology, Contact, Change*, pp. 23–41. Amsterdam/Philadelphia: John Benjamins.
- Nichols, J. (2003) ‘Diversity and Stability in Language’, in B. D. Joseph and R. D. Janda (eds) *The Handbook of Historical Linguistics*, pp. 283–310. Oxford: Blackwell.
- Oosterhoff, J. A. (2015) *Modern Dutch Grammar: A Practical Guide*. London: Routledge.
- Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., Pearse, W. (2013) ‘The Caper Package: Comparative Analysis of Phylogenetics and Evolution in R’, *R Package Version*, 5:1–36.
- Pagel, M. (1999) ‘Inferring the Historical Patterns of Biological Evolution’, *Nature*, 401:877–84.
- Pagel, M., Meade, A., and Barker, D. (2004) ‘Bayesian Estimation of Ancestral Character States on Phylogenies’, *Systematic Biology*, 53:673–84.
- Raftery, A. E. (1996) ‘Hypothesis Testing and Model Selection With Posterior Simulation’, in W. R. Gilks, D. Spiegelhalter, and S. Richardson (eds) *Practical Markov Chain Monte Carlo*, pp. 165–87. London, UK: Chapman and Hall.
- Revell, L. J. (2012) ‘Phytools: An R Package for Phylogenetic Comparative Biology (and Other Things)’, *Methods in Ecology and Evolution*, 3:217–23.
- Shcherbakova, O. et al. (2022) ‘A Quantitative Global Test of the Complexity Trade-Off Hypothesis: The Case of Nominal and Verbal Grammatical Marking,’ *Linguistics Vanguard*, 9:155–67.
- Shipley, B. (2000) ‘A New Inferential Test for Path Models Based on Directed Acyclic Graphs,’ *Structural Equation Modeling*, 7:206–18.
- Shipley, B. (2009) ‘Confirmatory Path Analysis in a Generalized Multilevel Context’, *Ecology*, 90:363–8.
- Shipley, B. (2016) *Cause and Correlation in Biology: A User’s Guide to Path Analysis, Structural Equations and Causal Inference*. Cambridge: CUP.
- Singer, R. (2016) *The Dynamics of Nominal Classification: Productive and Lexicalised Uses of Gender Agreement in Maung*. Number 642 in Pacific Linguistics. Boston: De Gruyter Mouton.
- Sinnemäki, K. (2019) ‘On the Distribution and Complexity of Gender and Numeral Classifiers’, in F. Di Garbo, B. Olsson, and B. Walchli (eds) *Grammatical Gender and Linguistic Complexity*, Vol. 2, pp. 133–200. Berlin, Germany: Language Science Press.
- Skirgård, H. et al. (2023) Grambank Reveals the Importance of Genealogical Constraints on Linguistic Diversity and Highlights the Impact of Language Loss. *Science Advances*, 9:eadg6175. <https://doi.org/10.1126/sciadv.adg6175>.
- Stolz, T., and Levkovich, N. (2022) ‘On the (Almost Im) Possible Emergence of Grammatical Gender in Language-Contact Situations,’ in N. Levkovich (ed.) *Susceptibility vs Resistance in Language-Contact Situations*, pp. 3–40. Berlin, Boston: Walter de Gruyter GmbH & Co KG.
- Trudgill, P. (2011) *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford, UK: OUP.
- van der Bijl, W. (2018) ‘Phylopath : Easy Phylogenetic Path Analysis in R’, *PeerJ*, 6:e4718.
- Veeman, H. et al. (2020) ‘Cross-Lingual Embeddings Reveal Universal and Lineage-Specific Patterns in Grammatical Gender Assignment’ in *Proceedings of the 24th Conference on Computational Natural Language Learning*, pp. 265–75. Association for Computational Linguistics. <https://aclanthology.org/volumes/2020.conll-1/>.
- Wichmann, S., Holman, E. W., and Brown, C. H. (2016) The ASJP Database (Version 17). Released on May 20 2016. <https://doi.org/10.5281/zenodo.3835942>.
- Xie, W. et al. (2011) ‘Improving Marginal Likelihood Estimation for Bayesian Phylogenetic Model Selection’, *Systematic Biology*, 60:150–60.