Gender comes first: Experimental evidence for the representation of gender closer to the noun stem than number across linguistic populations

Carmen Saldana[1,2,3], Patrick Kanampiu[4], and Jennifer Culbertson[4]
[1]Department of Translation and Language Sciences, Universitat Pompeu Fabra
[2]Department of Comparative Language Science, University of Zurich
[3]Center for the Interdisciplinary Study of Language Evolution , University of Zurich
[4]Center for Language Evolution, Univeristy of Edinburgh

Abstract

Languages exhibit a tremendous amount of variation in how they organise and order morphemes within words; however, regularities are also found. For example, gender and number inflectional morphology tend to appear together within a single affix. In the relatively rare cases in which they appear as two separate affixes, morphology encoding gender information tends to be placed closer to the stem than number morphology is. Alternative theories of gender and number have been designed (in part) to explain these tendencies. However, the typological data are very sparse, making it difficult to find clear evidence for the representations and mechanisms shaping the organisation of gender and number. Across a series of experiments, we use an artificial language learning paradigm to explore the inferences learners make about the order of gender and number affixes in the absence of any explicit information in the input. We show that language learners have a clear preference for placing nominal morphology containing gender information closest to the stem. This holds both for prefixing and suffixing morphology, and across different types of gender and number systems. Importantly, these biases are consistent across populations of participants with different L1 experience with gender morphology: English, Italian and Kĩĩtharaka speakers. These results provide a new source of empirical support for particular theories of gender, and in line with previous work, suggest that cognitive biases play a role in determining morpheme order in the world's languages.

Keywords: gender; noun class; number; morphology; morpheme order; typology; artificial language learning

1 Introduction

Languages differ what kinds of information encode in their morphology, how they encoded it in individual morphemes, and how those morphemes are organised into complex words. Despite this variation, one prominent class of regularities relates to how morphemes are ordered (Bybee, 1985; Greenberg, 1963). For example, derivational affixes (which change a stem's category or meaning) tend to appear closer to the stem than inflectional affixes (which do not). Similar trends are found for inflectional morphology. For example, when distinct affixes exist for number (e.g., plural) and case (e.g., accusative) and they appear together, number is ordered closer to the stem than case (Greenberg, 1963). While these observations have typically been made in the context of typological description or theoretical linguistics, there is a growing body of research which attempts to provide behavioural evidence to explain these kinds of ordering generalisations as the output of cognitive biases active during language learning (e.g., Maldonado, Saldana, & Culbertson, 2020; Mansfield et al., 2022; Saldana, Oseki, & Culbertson, 2021) and processing (e.g., Gibson et al., 2013; Hahn, Degen, & Futrell, 2021; Hawkins & Cutler, 1988; Hay, 2001). For example, previous research has shown that when people are trained on a novel artificial language that has separative case and number morphology, but no information about their relative order, they infer that number should be closer to the stem than case (Saldana et al., 2021). Here we apply this approach with the aim of better understanding the mechanisms which drive the linearisation of *gender* and number morphology. The interaction of gender and number is substantially more complex than number and case, and the typological data about how they are linearised is much more sparse. Nevertheless formal theories of gender and number make clear predictions, often based on this sparse data, about how humans represent these morphological categories. Below, we highlight aspects and issues in gender typology, discuss evidence for how gender and number are realised together, and outline theories and predictions about how learners might preferentially order them. We then report the results of artificial language learning experiments testing these predictions. To preview, we find that language learners have a clear preference for placing nominal morphology containing gender information closer to the stem than number information, thus revealing asymmetries between gender and number dependencies with the noun stem and suggesting that they are not processed at the same level in their realisation.

1.1 The typology of gender systems

Gender (or noun class)[1] is an inflectional feature found in many languages, which categorises nouns into two or more classes. By definition, gender is reflected not only on the noun but also in agreement morphology on elements external to the noun (e.g., determiners, verbs, auxiliaries or adjectives; Corbett, 1991). In languages with gender, various properties of nouns (and sometimes pronouns) often play a role in determining which gender class a noun falls into. While nouns can be assigned to gender classes arbitrarily, assignment can also be based on phonological or morphological features (e.g., French nouns that end in *on* are typically masculine gender; Italian nouns with the derivational suffix *-zione* are always feminine gender). Notably however, there is almost always a semantic core to gender assignment (see fig. 1 below); nouns in a given gender class will tend to show some overlap in their semantic features, even if normally *only* for a subset of them. This

[1] As is typical in mainstream typological literature, here we treat gender and noun classes as the same phenomenon. We use the term gender in this paper as a hyperonym for the two. We will thus use "gender class" instead of "noun class" to refer to a subset of nouns which inflect with the same gender morphology and thus the same agreement pattern.

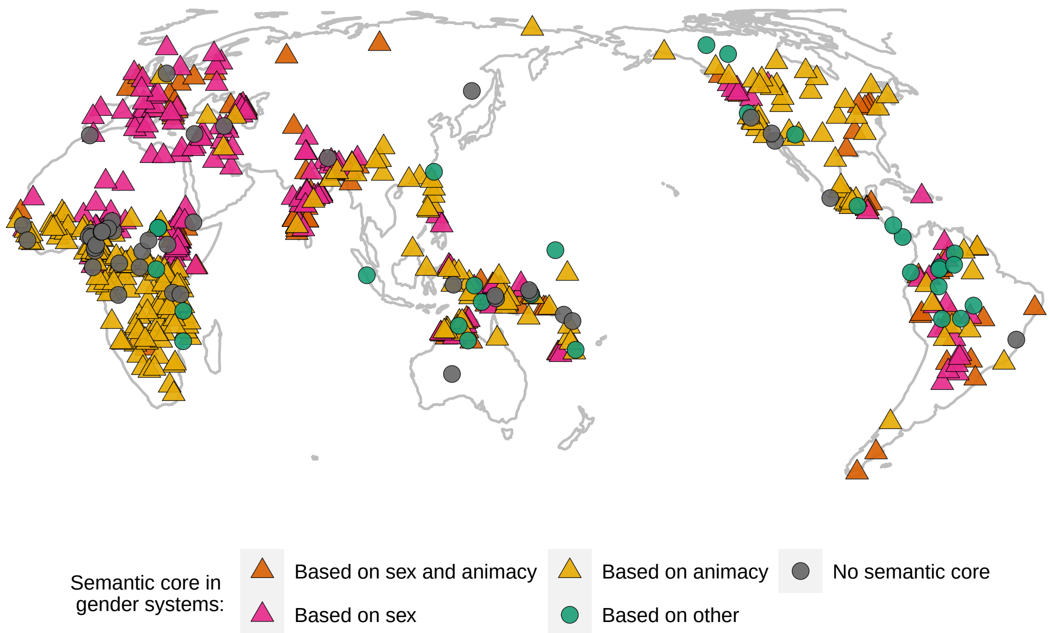


Figure 1

Distribution of different "semantic cores" in gender systems reported in Grambank (Skirgård et al., 2023). Of the 1,318 language in this dataset, 641 (48.6%) are (unequivocally) reported to have gender distinctions. A total of 93% of these gender systems have a documented semantic core; in 41.8% of the langauges (from 40 different language families) this core is animacy-based, in 26.7% (from 39 different language families) is sex-based, and in 19.7% (from 38 different language families) is both sex- and animacy-based; only 4.8% of the languages (from 20 different language families) have a semantic core based on some other property.[2]

subset of nouns (unlike the rest within the same gender class) can thus be considered to have gender assignment based on their semantics, that is, semantic gender assignment.

Semantic gender assignment within gender systems with a semantic core generally involves nouns denoting entities which are animate, or sex-differentiable; for example, animate entities might be assigned to one gender class and inanimates to another, or female entities to one gender class and males to another. In this paper we refer to the gender systems where animacy is a factor for class assignment in a subset of nouns as *animacy-based*, and to the systems where perceived gender or biological sex is a factor for class assignment in a subset of nouns as *sex-based* (though note, that they can co-exist; Corbett, 1991; Kramer, 2015). Figure 1 illustrates the high frequency of these systems. Examples of systems with animacy-based gender assignment can be found in

[2]Note that the smaller cross-linguistic sample in WALS (Corbett, 2013) suggested that sex-based was the majority semantic core; the larger Grambank dataset we use here (Skirgård et al., 2023) suggests this is not the case—animacy-based systems are more frequent.

Bantu languages (Atlantic-Congo) such as Kĩĩtharaka (see 1; Kanampiu, Martin, & Culbertson, under review). In Kĩĩtharaka, gender and number are indicated cumulatively in a single prefix (on nouns and also on agreeing elements). The class 1 (1a) morpheme, for example, is used for human entities in the singular, the class 2 (1b) morpheme is used for those same entities in the plural. By contrast, classes 3 (1c) and 4 (1d) are used for the singular and plural of specific inanimate entities.[3] Examples of systems with sex-based gender assignment can be found in Romance languages such as Italian (see 2-3). In Italian, entities denoting females are generally FEM, and males are generally MASC, but inanimate nouns are also categorised as MASC or FEM (e.g., 2a and 2c are both FEM).[4]

One can also find cases where the same noun stem can appear across gender classes, especially in sex-based systems. For example, in Italian, varying the gender of the noun stem indicating a particular type of animal denotes perceived biological gender: the same stem with *-a* or *-o* denotes the female or male entity respectively and will trigger the corresponding agreement (see 3). These nouns are often referred to as “common-gender” or “same root” nouns (Kramer, 2015). We will refer to this phenomenon as *variable* gender marking in opposition to *fixed* gender marking (e.g., 3).[5] While fixed gender marking can be assigned according to semantic properties or not, variable gender is *always* assigned according to semantics. Variable gender marking is rarer in non-sex-based gender systems.[6]

(1) *Kĩĩtharaka (Atlantic-Congo)*
Animacy-based system

a. mu-ntû
CL1-person
‘person’

b. a-ntû
CL2-person
‘people’

c. mû-tî
CL3-tree
‘tree’

d. mî-tî
CL4-tree
‘trees’

[3]In Kĩĩtharaka, as in other Bantu languages, the role of specific semantic features is somewhat contested, and morphophonological information, e.g., the nominal prefix often, but not always, correlates with gender.

[4]In Italian, as in other Romance languages, phonological features of inanimate nouns typically provide a cue to the noun’s gender. For example, in Italian most FEM nouns end in *-a* in the singular form, while most MASC nouns end in *-o*.

[5]There are theoretical accounts that consider variable gender marking of the Romance kind to be exclusively derivational (e.g., Mel’čuk, 2013), others which consider it equally inflectional as fixed marking (Ritter, 1993). Here we treat them both as inflectional as they pertain to the same gender marking system within s given language.

[6]In many Bantu languages there are productive class shifts which involve gender markers (Crisma, Marten, Sybesma, et al., 2011; Schadeberg, 2001). For example, in Kĩĩtharaka (see e.g., Kanampiu et al., under review) the stem in *mû-ana* ‘child’ (class 1) can take the prefix of another class to derive its diminutive form (i.e., *ka-ana* ‘small child’, class 12), or to derive its related quality or state (i.e., *(w)û-ana* ‘childhood’, class 14). These are sometimes analysed as derivational rather than inflectional due to the particular semantic relations involved (diminutive, augmentative, collective, manner, quality). Derivational gender markers are even sometimes added to the inflectional gender prefixes in Bantu languages (e.g., in Kikuyu *mû-raata* (CL1.friend, ‘friend’) → *ka-mû-raata* (CL12.CL1.friend), ‘small friend’); Stump, 1993), which suggests that they may be morphologically distinguished (Crisma et al., 2011; Di Garbo, 2014; Schadeberg, 2001). For this reason, these cases are not typically treated in theories of variable gender. However, see additional discussion in section 1.3.

(2) *Italian (Indo-European)*
Sex-based system, fixed gender examples

a. la barca
DEF.ART.FEM.SG boat.FEM.SG
'the boat'

b. le barche
DEF.ART.FEM.PL boat.FEM.PL
'the boats'

c. la madre
DEF.ART.FEM.SG mother.FEM.SG
'the mother'

d. le madri
DEF.ART.FEM.PL mother.FEM.PL
' the mothers'

(3) *Italian (Indo-European)*
Sex-based system, variable gender examples

a. la cerva
DEF.ART.FEM.SG deer.FEM.SG
'the (female) deer (sg)'

b. le cerve
DEF.ART.FEM.PL deer-FEM.PL
'the (female) deer (pl)'

c. il cervo
DEF.ART.MASC.SG deer-MASC.SG
'the deer (sg)'

d. i cervi
DEF.ART.MASC.PL deer-MASC.PL
'the deer (pl)'

1.2 The interaction of gender with number and their realisation

Gender appears to interact in a special way with number morphology. While there are many languages that have number but not gender, languages with gender tend to also have number (Corbett, 1991, 2000). Moreover, when both features are found in a language, gender and number are often marked via cumulative exponence (Corbett, 1991, 2000)—both features are marked with a single morpheme that is not further separable. For example, in Italian (see 2-3), noun suffixes all encode both gender *and* number (e.g., *-a* encodes FEM gender and SG number). Similarly, in Kîîtharaka (see 1), all prefixes encoding information about gender, also include number cumulatively. The cross-linguistic trend for gender and number morphology to be expressed cumulatively has been documented by Di Garbo (2014) in a sample of 100 languages (from 21 different families) spoken across Africa. She found that 81% percent of languages with gender have cumulative exponence of gender and number across all agreement targets, and 64% of languages that mark both gender and number on the noun also do it cumulatively.[7] This can also be seen in the AUTOTYP

[7] With the exception of the Afro-Asiatic Berber languages and the Atlantic-Congo language Wamey (which we discuss in the following paragraphs, see, e.g., 6), noncumulative or partially cumulative encodings tend to be found in languages where number marking on nouns is not obligatory but gender is.

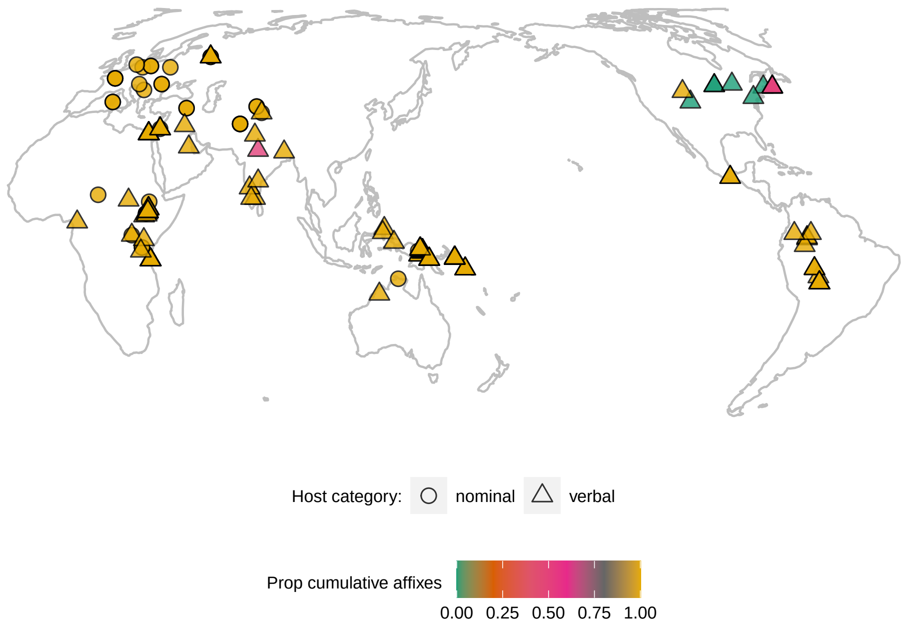


Figure 2

Proportion of cumulative exponence in gender-number affixes in the AUTOTYP grammatical markers data. The data shown includes 72 different languages from 24 different families. Triangles indicate verbal inflection systems and circles nominal inflection systems. Yellowish values indicate almost exclusively cumulative; greenish values indicate almost exclusively separative; orange and pink values indicate systems where there is a mix of both. Most languages in this sample tend to express gender and number cumulatively.

typological database (Bickel et al., 2023) as shown in fig. 2. We extracted information from 72 different languages (24 different families) with both gender and number. When gender is marked via affixation, it tends to appear within the same affix as number cumulatively, both on nominals (100% in this sample) and on verbs (88%).

Separative or partially separative expressions of gender and number are rarer. A prototypical example of what can be described as separative gender and number morphology is found in Spanish (Indo-european). Spanish has a gender system similar to Italian (see 4-5), but number morphology can be described separately from gender on both nouns and agreement targets. Feminine nouns are *typically* indicated by a final vowel *-a* (e.g., 4a), and masculine with *-o* (e.g., 4c). Plural is then marked with the additional word-final affix *-s* (e.g., 4b and 4d). The same separative morphemes are found on agreement targets. Spanish and closely related Romance languages spoken in the Iberian peninsula like Catalan, Galician and Portuguese are often used as an example case (see, e.g., Fuchs, Polinsky, & Scontras, 2015) to illustrate what is thought to be the typical *order* of gender and number morphology when they can be descriptively separated: gender tends to appears closer to the stem than number. While there are no sufficiently large cross-linguistic surveys of separative gender-number exponence to assess this tendency robustly, the available evidence suggests that this likely is the case (e.g., Di Garbo, 2014).

(4) *Spanish (Indo-European)*
Sex-based system, fixed gender examples:

a. la barca
DEF.ART.FEM.SG boat.FEM.SG
'the boat'

b. las barcas
DEF.ART.FEM.PL boat.FEM.PL
'the boats'

c. el faro
DEF.ART.MASC.SG lighthouse.MASC.SG
'the lighthouse'

d. los faros
DEF.ART.MASC.PL lighthouse.MASC.PL
'the lighthouses'

(5) *Spanish (Indo-European)*
Sex-based system, variable gender examples :

a. la gata
DEF.ART.FEM.SG cat.FEM.SG.
'the (female) cat'

b. las gatas
DEF.ART.FEM.PL cat.FEM.PL
'the (female) cats'

c. el gato
DEF.ART.MASC.SG cat.MASC.SG.
'the cat'

d. los gatos
DEF.ART.MASC.PL cat.MASC.PL
'the cats'

(6) *Wamey (Atlantic-Congo)*

a. a-san
SG.CL1-man
'man'

b. **v**e-san
PL.CL1-man
'men'

c. gu-meɗe
SG.CL15-night
'night'

d. **v**u-meɗe
PL.CL15-night
'nights'

(7) *Kinshasa Lingala (Atlantic-Congo)*

a. ba-to
CL2-person
'people'

b. ba-ma-loba
PL-CL6-word
'words'

Aside from these more straightforward examples, there are also cases in which gender information is encoded in some number morphemes but not others in a language. In these cases, there is further evidence of gender-closest orders. For example, in Wamey (Atlantic-Congo), the labial component common across most plural nouns classes *v-* or *w-* appears furthest away from the stem (see examples in 6; Jenkins & Jenkins, 2000; Santos, 1996). This labial segment has been analysed as a general plural marker which has been derived from what was originally the plural gender class 1 for human nouns (Santos, 1996). We find similar cases in other Atlantic-Congo families such as Bantu, where double plural marking can be used (allegedly with the same meaning as singleton plurals; Di Garbo, 2014). In these cases, a general plural marker (often a reanalysed human plural class marker) is placed furthest away from the stem, and the typical cumulative gender and number marker is placed closer. For example, in Kinshasa Lingala (Bokamba, 1977), the prefix *ba-*, typically used for plural human nouns (as in 7a) is argued to have been reanalysed as a general (redundant) plural marker. When used in this way, it precedes other gender-specific plural prefixes (as in 7b). In other words, the marker carrying gender information is closest to the noun stem. Similar examples of double plurals, where a class-independent general plural marker is placed further away from the stem than class-dependent plural marker can be found in Amharic (Kramer, 2016b). The opposite order—a general plural appearing closer to stem—has not been described for double plurals in the literature.

That said, there are potential counterexamples to the general trend for gender-closest order. For example, in Tachelhit Berber (Afro-Asiatic) MASC is unmarked, and FEM is marked with the prefix *t-*, which often appears as well as a *-t* suffix (as in 8a; Alderete, Jebbour, Kachoub, & Wilbee, 2015). For sex-differentiable referents, feminine nouns can be derived from masculine nouns by adding these *t* affixes (as in 8c). This can happen both in singular and plural forms. Plural MASC forms use a distinct initial vowel *i-* across both genders (as in 8d), as well as a plural suffix *-in* or *-an* for FEM and MASC genders respectively. In feminine plural nouns, we then observe the gender prefix *t-* (as in 8e) in the periphery, giving rise to a seemingly prefixal gender marker which is further away from the stem than the prefixal number marker (Alderete et al., 2015; Faust & Lahrouchi, 2022). Interestingly, this is not the case for suffixes. When the number suffix *-in* (PL.FEM) appears together with the FEM suffix *-t*, the number affix appears farther from the stem (as in 8f Faust & Lahrouchi, 2022). While similar systems are found in other Berber languages (Di Garbo, 2014), no other clear such examples have been documented, and even these appear to have gender-closest orders as well.

(8) *Tachelhit Berber (Afro-Asiatic)*

a. tafukt
FEM.sun.FEM

b. afruχ
boy

c. tafruχt
FEM.girl

d. ifrχan
PL.boy.MASC.PL.

e. tifrχin
FEM.PL.girl.FEM.PL.

f. tiʁrdatin
FEM.PL.mouse.FEM.PL.

To summarise, the typological evidence suggests that number and gender tend to be realised cumulatively. When they are expressed separately, it appears that gender is likely to appear closest to the stem; and in systems where the reverse order exists, gender-closest orders are also found. However, the data on separative morphology here are notably sparse. Both trends nonetheless have factored into formal theories of gender and number, where several distinct proposals have been made for how these features are formalised and mentally represented.

1.3 Theories of gender and number representation, and their impact on morpheme ordering

Formal linguistic theories differ substantially in terms of how gender is argued to be represented, and in how it interacts with number. These theories have been designed (although not uniquely) to account for the kind of cross-linguistic data discussed above, and make different predictions about the kinds of inferences human learners might make when learning gender and number systems. In relation to our object of study, the main source of divergence across these different theories is whether gender is representationally distinct from number in nominal morphology.

In one well-known class of theories, gender is represented—i.e., projected in the syntax—separately from number. Following Fuchs et al. (2015), we call these *split models*. Theories assum-

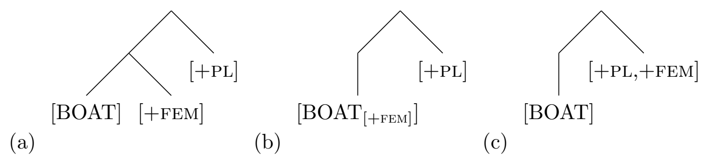


Figure 3

*Schematic illustrations of the representations posited by different models of gender and number: (a) split models with gender projecting outside the nominal, (b) lexicalist model with gender either specified in the lexical entry or represented as part of the nominal projected, (c) bundling model with gender represented together with number. In these theories, abstract features, like [+*FEM*] or meanings like [BOAT] are typically substituted for stems/morphemes either individually or in bundles from the mental lexicon at a subsequent level of representation (e.g., by the phonological system).*

ing a split between gender and number nevertheless diverge in whether or not gender is represented as part of the lexical entry of the noun. In one version of this model, illustrated in fig. 3(a), gender projects outside of the nominal head immediately below number(Antón-Méndez, Nicol, & Garrett, 2002; Carminati, 2005; Picallo, 1991).[8] Thus gender is local to the nominal head, but syntactically separate from it. However, intuitively, gender may seem to be an inherent property of a nominal head; in most cases, nouns have only one fixed gender, and gender can be assigned arbitrarily. This implies that gender may project in the syntax together with the nominal head, as in fig. 3(b). Indeed, there are proposals which argue that gender features are part of the lexical entry for a given noun. While these accounts capture the close relationship between nouns and gender, additional mechanisms are needed to capture systematic relationships between noun semantics (e.g., sex or animacy) and gender, and to deal with variable gender (i.e., stems that can be feminine or masculine depending on the properties of the referent). To account for these phenomena, lexicalist theories of gender often differentiate nouns that are specified for gender versus those which are unspecified (Alexiadou, 2004; Carstens et al., 2010; Harris, 1991; Roca, 1989).[9] Alternative theories which project the noun and gender features together use Distributed Morphology (Kihm, Cinque, & Kayne, 2005; Kramer, 2015; Lowenstamm, 2008). These accounts argue that nominal heads themselves have a more complex structure, with stems that are category neutral and a nominalising functional head, referred to as little *n*, where gender is located. Thus gender is within the nominal head and plays a role in turning a stem into a noun in a way that can be sensitive to semantics when necessary (see, e.g., Kihm et al., 2005; Kramer, 2015; Lowenstamm, 2008). There are also some split models which propose a distinction between variable and fixed gender nouns. For example, fixed gender might be projected within nP, and variable gender outside the nominal, within GenP (see Panagiotidis, 2018, for additional discussion of how this differs from a lexicalist model).

While split models (including lexicalist models) represent gender and number in an inherently asymmetrical way where gender is more local to the nominal head than number, an alternative model locates gender together with *number* in the NumP (i.e., either hosted on the number head

[8]In these theories gender heads its own syntactic projection, Gen(der)P(hrase), immediately dominating the n(oun)P(hrase), and dominated by the Num(ber)P(hrase) (De Belder & Van Koppen, 2016; Koopman, 2003; Picallo, 1991, however, cf. Picallo, 2008).

[9]For example, in some theories, nouns can have specified semantic properties or not. Nouns with a fixed semantic property (e.g., female or animate) are assigned gender via a lexical rule that connects semantics with grammatical gender. Nouns without a fixed semantic property (e.g., those with variable gender) are assigned gender via a lexical rule turns such nouns into pairs of nouns specified “male” and “female”, for example. Then a subsequent rule maps nouns denoting females to FEM gender, and the rest MASC by default (e.g., Carstens et al., 2010; Harris, 1991).

or expressed on the specifier within a number phrase; Carstens, 2003; Ritter, 1993).[10] Following Fuchs et al. (2015), we call this the *bundling model*. As with split models, within some bundling models fixed and variable gender is treated distinctly. For example, some bundling models argue that gender projects with number for variable (semantic) gender marking, but not for fixed gender, where it is considered to project with the nominal head (Carminati, 2005; De Vincenzi, 1999).

Split and bundling models capture different cross-linguistic tendencies regarding the interaction between gender and number morphology. If gender and number are bundled together and thus project together (as proposed by the bundling model), this predicts cumulative exponence, which is indeed extremely common. However, as discussed above, there are cases where the gender and number are expressed as separate morphemes. Moreover, there are many instances in which gender and number seem to act separately, even if they can be expressed cumulatively in other contexts. For example, Romance languages such as Italian and Spanish—which have both number and gender morphology—have verbal agreement with the subject noun in number but not gender. These phenomena (amongst others) suggest that at the very least gender and number *can* project separately. This possibility is captured by split models. Even though these models may differ with regard to whether the representation of gender is an inherent part of the noun or a projecting head immediately dominating it, they all predict that gender should be placed closer to the stem than number in cases where the exponence of gender and number is non-cumulative, a tendency tentatively supported by the cross-linguistic data discussed above. By contrast, bundling models make no clear predictions about order (i.e., without further ad-hoc ordering mechanisms forcing number to linearise outside of gender Ritter, 1993).

1.4 Other explanations of gender and number morpheme ordering

The theories discussed above make predictions (or not) about order on the basis of structural morphosyntactic representations of gender and number. However, there are other explanations of morpheme order that are potentially relevant for gender. For example, Greenberg (1963) introduced the notion of *proximity hierarchies* for nominal morphology, essentially dictating which morphemes are closer to the stem than others. He speculated that this was "related to degrees of logical and psychological remoteness" from the stem (Greenberg, 1963, , p. 103). In his proposal, proximity hierarchies were also related to other implicational hierarchies relevant to inflectional categories, for example, regarding markedness or syncretism. For case and number morphology, Greenberg (1963) placed number higher on the proximity hierarchy (i.e., closer to the stem) than case based on a set of observations: (i) number almost always appears closer to the stem than case cross-linguistically (Universal 39; Greenberg, 1963), (ii) many languages with systematic number marking lack case but not vice versa, and (iii) case is more likely to be neutralised within a given number value than number is to be neutralised within a given case value (Baerman, Brown, & Corbett, 2005; Corbett, 1991, 2000). Building on these notions, Bybee (1985) developed the *relevance* hierarchy. Under her proposal, morphemes which are more relevant to the stem will be placed closer to it. For case and number morphology, she suggests that number appears closer to the stem because it directly modifies the entity referred to by the noun (i.e., it is very relevant to the noun). By contrast, the case morpheme signals an external relationship between the entity and some event (i.e., less relevant to

[10] This is the case when gender is treated as an inflectional category; by contrast Ritter (1993) argues that gender is represented together with the noun if it is derivational (e.g., when a gender marker is used to form a diminutive of a noun typically in other class).

the noun stem), thus it appears further away from the stem. Neither Greenberg (1963)'s notion of proximity hierarchies nor Bybee (1985)'s relevance hierarchy has been explicitly applied to other nominal inflectional features. However, for gender and number morphology, the general line of argumentation would likely be less straightforward. Number should be placed higher in the proximity hierarchy than gender based on markedness or syncretism (e.g., there are many languages with the number category and without gender, but the reverse is rare; and gender is more likely to be the neutralised feature within a given number value and not vice versa Baerman et al., 2005; Corbett, 1991, 2000). However, as discussed above, gender is likely to be linearly closer to the stem when it appears separate form number (Di Garbo, 2014; Kramer, 2015), which would require gender to be higher on the proximity scale. Gender might be considered higher on the related relevance hierarchy as well, given that it can reflect highly specific semantic features of the stem, like sex and animacy.

More recently, alternative explanations of morpheme order focus on the role of distributional cues like the frequency with which stems and affixes (co-)occur. The general idea is that stem+affix pairs that co-occur more frequently together will be more likely to be ordered closer to one another (e.g., Hay, 2001; Hay & Plag, 2004; Ryan, 2010). In some cases, the relevant notion is degree of dependency, or strength of association, which can be quantified using mutual information between stem and affix. Hahn et al. (2021) show that this is a good predictor of the relative order of verbal affixes in a number of different language. Rathi, Hahn, and Futrell (2022) further argue that this same measure can predict patterns of fusion between morphemes (i.e., the degree to which formatives are phonologically intertwined with their host). This is reminiscent of accounts of word order arguing that mutual information can predict linear distance between head and dependents (Culbertson, Schouwstra, & Kirby, 2020; Futrell, Dyer, & Scontras, 2020; Futrell, Mahowald, & Gibson, 2015; Hahn, Jurafsky, & Futrell, 2020). Distributional statistics like these have the potential to make predictions about gender and number. Notably, stems and gender affixes are likely to co-occur very frequently, more so than stems and number markers. For example, in Spanish, where number and gender are not cumulative, singular is zero-marked while gender is always overtly marked. In this case, the co-occurrence frequency (and mutual information) between a stem and a gender morpheme is likely to be higher than between a stem and a number morpheme.[11]

1.5 Experimental evidence for preferences in morpheme ordering

The pervasiveness of cumulative exponence of gender-number morphology and scarcity of documented separative exponence makes it difficult to assess whether there is support for theories which predict clear ordering preferences among these two inflectional categories. Moreover, the frequent coexistence of fixed and variable gender within the same system of gender marking make it difficult to explore potential differences in their linguistic representations. However, even if natural language data were plentiful and typological regularities were robustly quantified, cross-linguistic tendencies are not necessarily directly linked to individual-level mental representations or biases. Tendencies could reflect historical contingencies, cognition-external pathways of change, or even common distributional properties of the linguistic input that reflect communicative needs rather than

[11] In Spanish at least, this is equally the case for both fixed and variable gender. However, as noted above, some accounts of order focus on the relationships between concepts expressed by words and morphemes (Culbertson et al., 2020), rather than co-occurrences between particular stems and markers. This might suggest an asymmetry between variable gender and fixed gender referents: the association between a particular concept and it semantic features may be stronger for fixed compared to variable gender referents. We leave this possibility aside here, but will come back to it in the General Discussion.

learning biases or mental representations. For instance, as we have summarised in the previous section, some studies have shown that co-occurrence statistics among stems and morphemes, rather than any abstract representations of their structure, may determine morpheme order and morphological fusion both within and across languages (Hahn et al., 2021; Hahn, Mathew, & Degen, 2022; Hay, 2001; Hay & Plag, 2004; Rathi et al., 2022).

Experimental research offers a way to test these different theories, and in particular to uncover evidence for representations or biases that shape gender and number systems. For example, if particular ways of ordering morphemes are preferred by learners in carefully controlled experiments, this can provide evidence for specific theories or representations. It also provides a potential link between human cognitive biases and typology: individual learners' preferences can be amplified as languages are transmitted from generation to generation (Kirby, Tamariz, Cornish, & Smith, 2015; Thompson, Kirby, & Smith, 2016). In recent work, Saldana et al. (2021) used a series of artificial language learning studies to show that English- and Japanese-speaking participants' inferences about number and case morphemes follow the cross-linguistic trend in having number closer to a noun stem than case. This held even though co-occurrence statistics among stems and different affixes were kept constant. Saldana et al. (2021) argue that this ordering preference therefore likely reflects differences in the kinds of meanings conveyed by these two types of morphemes (e.g., following Bybee, 1985; Culbertson et al., 2020).

Here, we use a similar experimental design to probe the biases of language learners with regards to the relative order of gender and number morphology. We train participants on nominal systems with separative number and gender morphology, but holding out any evidence about the relative order of these morphemes in the language. By holding distributional information—that is, co-occurrence frequencies between stems and morphemes—constant, we can isolate the impact of the features themselves. For example, we can directly test whether learners' biases more closely align with split models of gender and number, or bundling models. As discussed above, split models predict that gender should be linearised closer to the noun than number. By contrast, if gender and number are bundled together, we wouldn't predict any particular ordering pattern to be more likely than another (without further constraints). If these hypothesised structural representations in fact play no role, and order is fully determined by distributional information accessible to the learner during acquisition, then equally, we should see no preference for one order over the other when input frequencies are controlled. In the experiments described below we also explore whether learners' ordering preferences are consistent regardless of experience with gender systems in their L1, *and* whether preferences differ when participants are taught fixed and variable gender systems. As mentioned above, in some formal theories, variable and fixed gender have been argued to behave in different ways (with variable gender projecting higher, either with number or with its own head Carminati, 2005; De Vincenzi, 1999; Panagiotidis, 2018), which could potentially impact the strength of any ordering preferences—e.g., weaker for variable gender systems than for fixed gender systems.

2 Study 1: ordering preferences for gender and number morphemes across populations

In the first set of experiments, we train English-, Italian-, and Kĩĩtharaka-speaking adult participants on miniature nominal paradigms with separative exponence of gender and number morphology. For the sake of simplicity and learnability, we focus on nominal gender marking (and not on agreement targets), and gender assignment is exclusively (and not only partly) based on the semantics of the nouns. Gender assignment can be either based on animacy and fixed (i.e., stems denote either animate or inanimate entities), or based on sex and variable (i.e., stems denote only sex-differentiable entities, and they can be either male or female). We refer to these conditions as *animacy-based* and *sex-based* respectively.

Participants' training input indicates whether gender and number affixes generally precede or follow the noun stem (i.e., they are either all prefixes or all suffixes), but crucially participants are not given any examples in which overt gender and number affixes co-occur on the same noun. Participants thus only know whether the novel language is prefixal or suffixal—gender and number marking can be either both prefixes or both suffixes—but they have no evidence indicating in which way gender and number affixes are ordered when they appear together—both as affixes or both as prefixes—within the same noun. At test, they are asked to produce an order for these held out examples. These productions will reveal the inferences learners make regarding the relative order of gender and number morphology, in the absence of any evidence in the input linguistic system. We compare across speakers of different languages in order to assess whether substantially different experience with gender systems in their first language impacts the results.

English-speaking participants have an L1 that marks (plural) number on nouns, but does not mark gender.[12] Thus these participants have no evidence from their native language (neither for sex-based nor for animacy-based systems) of how inflectional morphemes of gender and number should be ordered relative to one another. They do, however, have evidence of derivational gender morphemes being placed closer to the root than (inflectional) number (e.g., *lion-ess-(e)s*), which in the absence of any other gender linguistic representation, could lead to participants placing sex-based gender markers closer to the stem because they treat them as derivational morphology. The animacy-based system we use here is quite distinct from this. Further, in contrast to English, Italian-speaking participants have an L1 which marks both gender and number, but cumulatively (see2-3 above). The system can be characterised as sex-based, with both variable and fixed sub-systems. If Italian speakers also prefer to place number in the periphery, it will not be because they are only familiar with inflectional affixes with number information alone (see Saldana et al., 2021, for related evidence). Their knowledge of gender also makes it less likely that they would fall back on experience with derivational morphology to determine order. Finally, Kîîtharaka-speaking participants have an L1 that is like Italian in featuring cumulative gender and number. However, like other Bantu languages, it is prefixing. Similar to Kîîtharaka, described above, the Kîîtharaka gender system can be described as containing an animacy-based semantic core. It has a relatively large number of classes, typically treated as singular-plural pairs, with classes partly determined by semantics. The most obviously inflectional semantic feature is human vs. non-human (e.g., CL1 and CL2 nouns are exclusively human, singular and plural respectively; Kanampiu et al., under review). Some Kîîtharaka class markers have been analysed as derivational (since they can be used to express diminutive, augmentative, pejorative meanings). However, unlike English and Italian, Kîîtharaka does not have any derivational morphology that expresses sex differences, and derivational uses of gender morphemes still cumulatively express number.[13]

To preview, the order participants across all three languages infer indicates a clear preference for placing gender closest to the noun, in accordance with a representation in which gender is more closely tied to the noun than number is. This is straightforwardly predicted by split models

[12] English is not considered to have a (grammatical) gender system. English does not mark gender on the noun and there is no further agreement targets within the language other than on third person singular personal pronouns.

[13] As noted above for other Bantu languages, in some cases cumulative derivational-number morphemes can be stacked on top of cumulative gender-number morphemes.

(e.g., Harris, 1991; Kramer, 2015; Picallo, 1991) which argue that gender and number are projected separately, with gender either part of the nominal project or immediately dominating it. By contrast, to account for these results, bundling models would require ad-hoc ordering mechanisms (e.g., Ritter, 1993). We also find that Italian speakers are slightly more likely to produce the reverse orders for variable gender systems than for fixed gender systems. This is not the case for English speakers, where we find no difference between gender systems. These results therefore expand the range of behavioural evidence for the role of cognitive representations in determining morpheme order, and further suggest a possible difference between the representations of gender across variable and fixed systems (consistent with the models proposed in , e.g., Carminati, 2005; De Vincenzi, 1999) worth exploring further in future work.

2.1 Materials and methods

Our experiments use an extrapolation design to test participants' preferences for gender and number morpheme ordering on nouns (following Saldana et al., 2021). Participants were trained on a subset of a nominal system with two gender classes and two number values (singular and plural). Across all conditions, only one gender class is overtly marked via affixation, the other is not; and only one number value is overtly marked via affixation, the other is not. The system is thus designed such that only some forms involve affixes of both gender *and* number. Crucially, these forms are held out during training, but participants are asked to infer them during testing. During training, participants therefore learn only how *single* morphemes are ordered relative to the noun stem. At test, they must therefore infer the relative order of the two. They can either place the gender morpheme closer to the noun stem than the number morpheme, or vice versa.

Experiments with English and Italian speakers were designed using a 2×2 between-subjects factorial design where we manipulated the position of affixes (either all suffixes or all prefixes) and the type of gender system (either animacy-based or sex-based). In one condition, the gender systems mark perceived biological sex of nonhuman animals with sexual dimorphism. For example, the stem alone is used to refer to the animal with female characteristics (FEM), and the stem together with an overt gender affix is used when the same denoted animal shows male characteristics (MASC). We refer to this condition as *sex-based*, but it also corresponds to what we have called a variable system above, since the *same* stem encoding a single meaning (type of animal) can be inflected for both genders (here via an overt affix vs zero-marking). The other condition, which we call *animacy-based*, is closer to what we have called a fixed gender system. In this condition, gender distinguishes between animate (ANIM) and inanimate (INANIM) entities and the denoted entity type cannot be both ANIM and INANIM. Stems denoting animates are marked with an affix, and stems denoting inanimates are zero-marked.[14]. We want to control for stem-affix co-occurrence statistics equally across conditions and we need to share the *stem forms* across classes in fixed animacy-based systems as well as in variable sex-based systems; the only way to do it by having homophonous stems across classes in the animacy-based condition. The correspondence between the animate and inanimate meanings of two homophonous stems is therefore arbitrary in the fixed

[14]Note that for English-speaking participants, we had initially preregistered to manipulate whether the overt vs. unmarked meaning was female or male (in the sex-based system), and animate or inanimate (in the animacy-based system). This made no difference to the results, and therefore we did not plan to perform this kind of manipulation with our Italian population (see preregistration at osf.io/Italian). In this paper we only focus on the English data comparable to Italian (with female and animate values marked); however, the results for all the manipulations initially preregistered with English speakers are available in the analysis script in osf.io/genderNumber.

animacy-based systems (see 1). Having shared stem forms in both sex-based and animacy-based conditions allows us to match co-occurrence frequencies (and mutual information) between each combination of stem form and affix. Thus learners exposed to *both* variable and fixed systems must make inferences about order in the absence of distributional information.

The design used for Kĩĩtharaka speakers diverges from this. After piloting, we determined that these systems were not readily learned by our Kĩĩtharaka population (described in detail below), therefore we went with a simpler system that is more natural-language-like.[15] Firstly, gender classes convey a more salient animacy distinction for the target population, that is, between humans and inanimate objects. Secondly, in this system, stems are not shared across gender classes, but are unique–as they are mostly in natural language. Note then that here the co-occurrence frequencies of stem+affix combinations are *not* matched; the same stem form can only appear in a single gender class, and thus stems overtly marked for gender are always marked for gender (and never appear with zero marking). We discuss the implications of this difference in light of the results below.

2.1.1 Artificial lexicon

Participants were trained on a language with a small lexicon of four (English and Italian participants) or eight (Kĩĩtharaka participants) stem forms[16], and two affixes, one expressing number (singular or plural) and the other expressing gender (sex-based or animacy-based assignment depending on the condition). Across all conditions, singular was always unmarked and plural was marked; similarly, one gender value was always marked with an affix and the other is not. For English and Italian speakers, in sex-based conditions MASC was overtly marked and FEM unmarked (i.e., we ensure that the gender marker is not as straightforwardly mapped to the derivational feminine gender affix in the speakers' native languages); in animacy-based system, ANIM was marked and INANIM unmarked. For Kĩĩtharaka participants INANIM was always unmarked as well. The two affixes were randomly chosen by-participant from the set {gu, sa, vi} (English), {gu, sa, di} (Italian), or {ke, bu, ra} (Kĩĩtharaka).[17] Gender and number affixes were attached to stems, with the relative order of affixes and nouns (prefixing or suffixing) determined by the condition. For English and Italian participants, we tested both suffixing and prefixing, for Kĩĩtharaka participants we tested only prefixing. Examples of a full lexicon for each gender system type and each language population are shown in Table 1.[18]

During training phases (described below), participants saw three kinds of nominal forms in the language: the stem alone (singular, unmarked class), gender-marked stems (singular, marked

[15] We are not entirely sure why this is case. It could be that learning gender as distinct from number is particular difficult for speakers of Kĩĩtharaka due to the way their native language works. Alternatively, it could be that the population we test is simply not as experienced with taking part in experimental studies compared to our English- and Italian-speaking participants, who were recruited via Prolific. It's worth noting that also they struggled to learn the morphemes and their meanings in our piloting, we never observed a preference for number closer to noun than gender.

[16] Note that participants are not required to learn the stems per se. As described below, in testing phases, the correct stem is always provided.

[17] These sets of affixes were modified across the different linguistic populations to fit the participants' L1 phonotactics and avoid already existing words.

[18] In several respects, the system we chose to use for Kĩĩtharaka speakers is maximally similar to Kĩĩtharaka: it is prefixal, the major inflectionally-relevant semantic distinction for gender class is human-ness, and humans are marked relative to inanimate objects. The latter is not strictly-speaking the case for Kĩĩtharaka, but, for example, the main class including humans uses on overt prefix, and the most heterogeneous class, into which most loan words fall, uses a null prefix. We made this choices because we had access to only a limited number of participants and, as noted above, piloting indicated that participants struggled to acquire the morphemes.

class), and number-marked stems (plural, unmarked class). Crucially, forms in which both a number and gender overt morphology would be required—i.e., plural, gender-marked class—were held out. As noted above, for English and Italian, each of the four stems is used equally frequently in both gender classes, and with both numbers, therefore no co-occurrence statistics that differentiate stem+number from stem+gender combinations are present during training. In Kîîtharaka, because stems are not shared across classes, marked gender class stems always occurred with the gender marker. Thus co-occurrence statistics would indicate to participants that stem+gender combinations are more frequent.

2.1.2 Experimental procedure

The experiments were programmed using *JsPsych* (De Leeuw, 2015) and displayed in participants' browser window. Before starting the experiment proper, participants were told they would learn how to describe simple pictures in a new language. The training phase was divided into three sub-phases: exposure, picture-matching, and recall. During exposure, participants saw images along with orthographically presented nouns using a stem alone, or a stem with a single affix (either number or gender). More specifically, they were trained on three different types of inflected forms (see Table 1): the stem alone (G1.SG, where SG $\rightarrow \emptyset$ and G1 $\rightarrow \emptyset$), gender marked stems (G2.SG, where SG $\rightarrow \emptyset$ and G2 $\rightarrow$ affix), and number marked stems (G1.PL, where PL $\rightarrow$ affix and G1 $\rightarrow \emptyset$). Participants saw each stem in each of these three forms five times, for a total of 60 trials (randomly ordered for each participant). In the picture-matching phase, participants saw two images and a form and had to click the corresponding image. As in the exposure phase, forms featured either a stem alone, or a stem with a single morpheme (either number of gender). The foil image for each trial was either the wrong gender, or the wrong number. Participants saw each stem with each pairing of correct and foil three times, for a total of 48 matching trials (randomly ordered for each participant). In the recall phase, participants were tested on the forms they had been trained on so far (i.e., G1.SG, G1.PL and G2.SG). In each trial, they saw an image and had to construct the corresponding form by clicking buttons. The button set always included the correct stem, the gender affix, and the number affix, but these were randomly ordered on each trial. Full feedback was provided. Participants saw each stem inflected for G1.SG, G1.PL and G2.SG three times (36 trials in total, randomly ordered for each participant).

In the critical testing phase, participants were tested on the held-out G2.PL form-meaning mappings: where both G2 gender and PL number are overtly marked, and therefore they need to provide forms with the two affixes. Participants saw an image and had to construct an inflected form by clicking buttons. The button set always included the correct stem, the marked gender affix (depending on the condition), and the number PL affix, but these were randomly ordered on each trial. Participants had to use all three buttons and could not submit their answer until they did. No other feedback was provided. Participants saw each inflected stem three times (total of 12 trials, randomly ordered for each participant).

2.1.3 Participants

English and Italian-speaking participants were recruited through Prolific (www.prolific.co) and paid 10 GBP/h for participation in the study, which took approximately 12 minutes. We used Prolific screening criteria to include native English- and Italian-speaking participants who were

English (sex-based)			English (animacy-based)		
SG		PL	SG		PL
G1:FEM	G2:MASC	G1:FEM	G1:INANIM	G2:ANIM	G1:INANIM
dur-∅	dur-*sa*	dur-*gu*	shib-∅	shib-*sa*	shib-*gu*
lan-∅	lan-*sa*	lan-*gu*	kot-∅	kot-*sa*	kot-*gu*
pek-∅	pek-*sa*	pek-*gu*	weil-∅	weil-*sa*	weil-*gu*
chit-∅	chit-*sa*	chit-*gu*	houf-∅	houf-*sa*	houf-*gu*

(a) *English example paradigms.*

Italian (sex-based)			Italian (animacy-based)		
SG		PL	SG		PL
G1:FEM	G2:MASC	G1:FEM	G1:INANIM	G2:ANIM	G1:INANIM
cheru-∅	cheru-*sa*	cheru-*gu*	peta-∅	peta-*sa*	peta-*gu*
lonu-∅	lonu-*sa*	lonu-*gu*	gaze-∅	gaze-*sa*	gaze-*gu*
povu-∅	povu-*sa*	povu-*gu*	balo-∅	balo-*sa*	balo-*gu*
kalu-∅	kalu-*sa*	kalu-*gu*	cavu-∅	cavu-*sa*	cavu-*gu*

(b) *Italian example paradigms.*

Kîîtharaka (animacy-based)		
SG		PL
G1:INANIM	G2:HUMAN	G1:INANIM
∅-caanu	*ra*-ariû	*bu*-caanu
∅-buri	*ra*-ene	*bu*-buri
∅-gogo	*ra*-rui	*bu*-gogo
∅-îtî	*ra*-ciarî	*bu*-îtî

(c) *Kîîtharaka example paradigm.*

Table 1

Example training artificial lexicons for English (a), Italian (b) suffixal gender systems, and Kîîtharaka prefixal system (c) in Study 1 experiments. During training, participants saw the zero-marked gender forms both in singular and plural, but marked forms only in singular. During the testing phase, participants had to produce the held-out forms, requiring gender and number (plural) marking.

raised monolingual and whose primary language is still their native language.[19]

[19] English materials, including artificial lexicon, experiment instructions, and participant information and consent document, were created by the third author, a native-English speaker. Italian materials were translated or revised as necessary

Kîîtharaka-speaking participants were recruited via local contacts in Kenya (specifically, in Tharaka-Nithi county, the main region where the language is spoken). Participants were largely students of Tharaka University, most of whom were multilingual speakers of Swahili, English or both. Participants were sent a link to the experiment, and after completing the experiment, they were paid 400 KES.

As per our preregistration, participants who got less than 75% correct on recall trials in the training phase were excluded (N = 9 both for English and for Italian, and N = 1 for Kîîtharaka).[20] For English and Italian speakers, a total of 80 participants' data were analysed for each language (20 in each condition of our 2*x*2 factorial design). For Kîîtharaka speakers, a total of 28 participants' data was analysed (in a single condition, as described above).

2.1.4 Data analysis

For each population tested, we used *R*'s *brms* (Bürkner, 2018) library as an interface to *Stan* (Carpenter et al., 2017) to fit a mixed-effects Bayesian binomial model predicting participants' production of gender-closest order (stem-G2-PL if suffixing or PL-G2-stem if prefixing). Our dependent variable was participants' responses for each of the critical test trials (coded as 1 if gender-closest, and 0 if number-closest). For English and Italian, we included fixed effects of affix position (suffixing vs prefixing, treatment-coded) and gender system (sex-based vs animacy-based, treatment-coded), as well as their interaction. As random effects, we included intercepts for participants as well as stems. We set the same Student-*t* prior on all fixed effects and intercepts ($DF = 6, \mu = 0, \sigma = 1.5$); for random effects' standard deviations, we set a half-Cauchy prior with a scale parameter 1. The model for the experiment with Kîîtharaka speakers was intercept-only (as we only have one condition); the same random effects structure and priors were set as in the previous models.

We ran two further models, one comparing the experiments with English and Italian speakers and another one comparing the prefix animacy-based conditions across all three linguistic populations (Kîîtharaka, English and Italian). In the former model we used the same model structure as the one described for English and Italian above with the added fixed effect of language group (English vs Italian) and an interaction term. In the latter model, we used the same model as for the experiment of Kîîtharaka speakers alone but instead of an intercept-only model, we add the effect of linguistic population (Kîîtharaka, English and Italian). We used Kîîtharaka as the baseline, since the artificial system was designed to mirror closely that of Kîîtharaka—with the only difference being that morphology was separative—and the co-occurrence statistics in the experiment with Kîîtharaka speakers were much stronger between stems and gender affixes than between stems and number affixes.

2.1.5 Transparency and openness

All data and analyses reported are available at osf.io/genderNumber. The preregistered hypothesis, design and analysis plan can be found in osf.io/English (for English) and osf.io/Italian (for Italian). We preregistered English and Italian language groups as different experiments, and the

by the first author, a fluent Italian speaker, and checked by a native speaker. Kîîtharaka materials were translated or revised as necessary by the second author, a native Kîîtharaka speaker.

[20] Note that we further excluded trials which did not include the two affixes in the response or that did not place both affixes before or after the stem (N=14 trials in the English group, and N=13 trials in the Italian group).

latter as a replication of the former with a different language group.[21] The Kîîtharaka experiment was not preregistered.

2.2 Results

Based on our preregistered hypothesis, we predicted that participants would prefer to place gender morphology closer to the stem than number morphology in systems where both gender and number affixes are separate and both are placed before or after the stem. In the experiments with English and Italian speakers, we further explore whether participants are equally likely to produce gender-closest orders across animacy-based (fixed) and sex-based (variable) gender systems. As a reminder, in variable systems lexical stems map to a single meaning and are inflected for different gender feature values (as in the sex-based system in our experiments); in fixed gender systems lexical stems are either shared idiosyncratically across classes (as in the animacy-based systems in our English and Italian experiments), or are not shared at all (as in the animacy-based system in our Kîîtharaka experiment).

Figure 4 shows average proportions of gender-closest orders in the critical trials for participants in the populations and conditions we tested. A visual inspection suggests that our results are consistent with the preregistered predictions: participants infer gender-closest orders (i.e., stem-G2-PL or PL-G2-stem), regardless of affix position, the type of gender system, or the participant population.

The model results confirm this. For English-speaking participants, the intercept suggests very strong evidence that English speakers prefer to produce gender-closest as opposed to number-closest orders in animacy-based prefixal systems ($\hat{\beta} = 6.348$, 90%CI = $[4.779, 8.189]$, $P(\hat{\beta} > 0) = 1$). This preference is not weaker in other conditions (sex-based vs animacy-based prefixal: $\hat{\beta} = -0.857$, 90%CI = $[-2.426, 0.661]$, $P(\hat{\beta} > 0) = 0.821$; interaction with suffixal systems: $\hat{\beta} = -0.300$, 90%CI = $[-1.969, 1.304]$, $P(\hat{\beta} < 0) = 0.618$; animacy-based suffixal vs prefixal: $\hat{\beta} = -0.721$, 90%CI = $[-2.179, 0.686]$, $P(\hat{\beta} < 0) = 0.802$).

For Italian-speaking participants, the model also confirms notably above-chance production of gender-closest orders in the reference animacy-based prefixal condition ($\hat{\beta} = 7.235$, 90%CI = $[5.373, 9.434]$, $P(\hat{\beta} > 0) = 1$), which is equally strong for animacy-based suffixal systems ($\hat{\beta} = -0.550$, 90%CI = $[-2.217, 1.087]$, $P(\hat{\beta} > 0) = 0.709$). We find evidence (although weak, $P(\hat{\beta} < 0) < 0.95$) suggesting that Italian speakers are slightly less likely to produce gender-closest orders in sex-based systems (sex-based vs animacy-based prefixal: $\hat{\beta} = -1.628$, 90%CI = $[-3.489, 0.094]$, $P(\hat{\beta} < 0) = 0.941$; interaction with suffixal order: $\hat{\beta} = -0.523$, 90%CI = $[-2.416, 1.316]$, $P(\hat{\beta} < 0) = 0.676$). We observed this same tendency with the English data, however,in that case the evidence was not decisive ($P(\hat{\beta} < 0) \approx 0.8$). However, in a further model combining the data from English and Italian speakers we find the same effect of system type in the data from English speakers (sex-based $<$ animacy-based) and no difference with the Italian-speakers' data (sex-based vs animacy-based prefixal with English speakers: $\hat{\beta} = -1.315$, 90%CI = $[-2.899, 0.210]$, $P(\hat{\beta} < 0) = 0.923$; interaction with language group: $\hat{\beta} = -0.887$, 90%CI = $[-2.606, 0.773]$, $P(\hat{\beta} < 0) = 0.808$). Note however, that the evidence is weak across models

[21] As noted above, the original experiment with English speakers counter-balanced which gender class was marked and unmarked; however the preference towards gender-closest orders was found across the board, and we only include the data of English participants for the same conditions as those run in Italian. We nevertheless provide these excluded data in the supplementary materials available at osf.io/genderNumber.

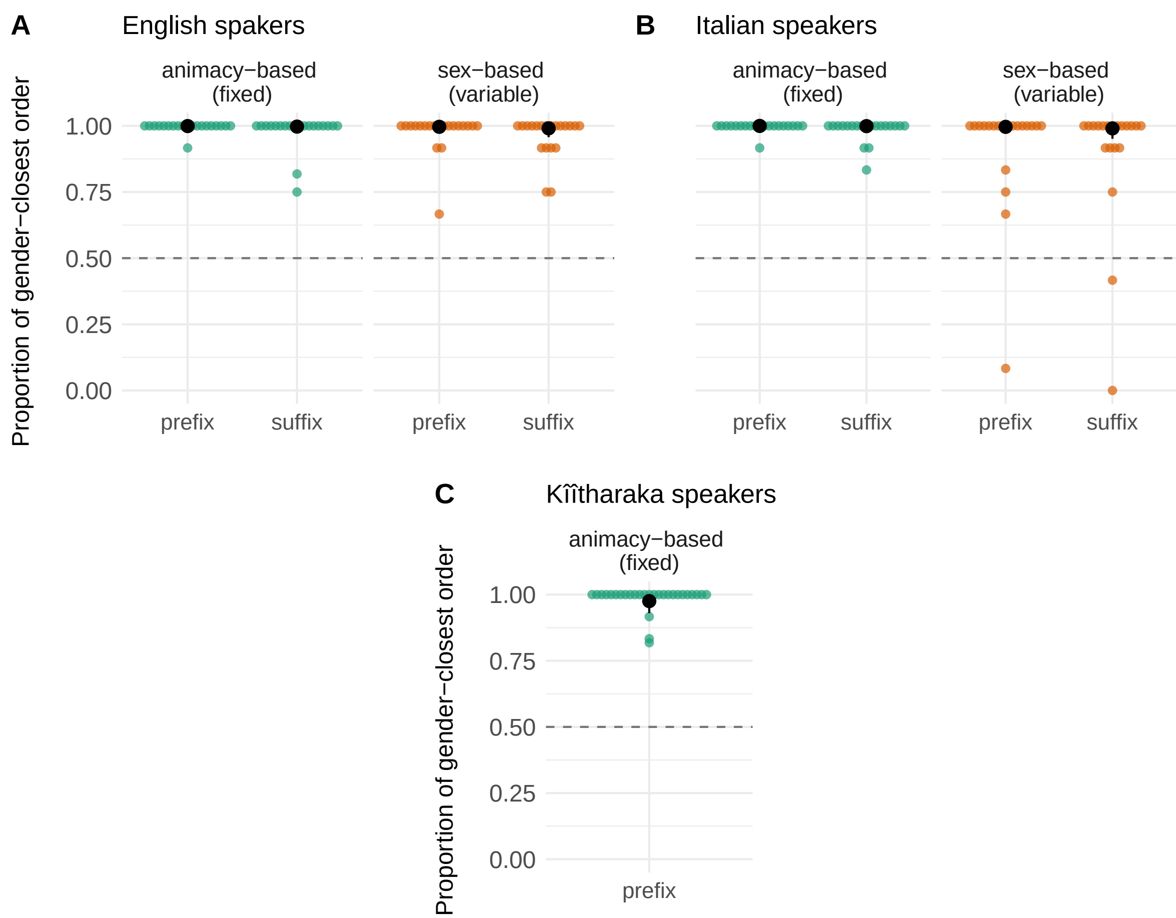


Figure 4

Average proportion of responses with gender-closest order among English- (A), Italian-speaking (B), and Kĩĩtharaka-speaking participants (C) in Study 1, by affix position and gender system type. Coloured dots represent participants' individual scores; black dots represent the model's predicted mean accuracy scores and the error bars represent the model's predicted 90% credible intervals. Dashed lines represent the chance level.

($P(\hat{\beta} < 0) < 0.95$) and it could be driven by the stark homogeneity in our data, where almost everyone chooses gender-closest orders categorically and a few outliers can easily drive differences across conditions.

Finally, and unsurprisingly, for Kĩĩtharaka-speaking participants, the model intercept suggests very strong evidence for a preference to produce gender-closest as opposed to number-closest orders ($\hat{\beta} = 4.623$, 90%CI = $[2.590, 6.751]$, $P(\hat{\beta} > 0) = 0.993$). Interestingly however, a further model comparing the accuracy in the animacy-based prefixal condition across language groups suggests that this preference for gender-closest orders is comparable (Italian vs Kĩĩtharaka: $\hat{\beta} = 0.674$, 90%CI = $[-0.870, 2.312]$, $P(\hat{\beta} > 0) = 0.758$; English vs Kĩĩtharaka: $\hat{\beta} = 0.670$, 90%CI = $[-0.878, 2.373]$, $P(\hat{\beta} < 0) = 0.753$), thus suggesting that the preference for gender-closest orders is not aided by very strong co-occurrence statistics (as we find in the system used in the experiment with Kĩĩtharaka speakers).

2.3 Discussion

The set of experiments in Study 1 tested whether language learners with different native languages would infer that a novel gender morpheme should be ordered closer to the noun stem, with a novel number morpheme placed peripherally. As described above, the precise systems participants were taught differed across populations. Importantly, the training input we used for English and Italian speakers was designed so that no co-occurrence statistics would lead learners to prefer one order over the other (suggested as a mechanism for ordering preferences by, e.g., Hahn et al., 2022). In other words, stems occurred equally frequently with the gender and number morpheme. This was not the case for Kĩĩtharaka-speakings participants, for whom we used a simplified language that resembled typical natural language gender systems: stems marked with an overt gender affix *always* occurred with that affix. Nevertheless, our results confirm a strong preference for oredring gender morphemes closer to the stem than number morphemes across all populations. The fact that the preference was equally strong across these populations suggests that higher co-occurrence frequencies are not necessary to skew participants' performance in favour of gender-closest order. The similar strength of preference regardless of native language experience with gender further suggests that the preferences are not driven by knowledge of the relative order of derivational and inflectional affixes. While this could potentially explain English speakers' placement of gender closer to the stem—i.e., if their lack of experience with gender leads them to treat this morpheme as derivational—it is not likely to explain the results for Italian or Kĩĩtharaka speakers. Italian speakers arguably don't have any reason to treat the systems they are trained on here as anything other than akin to gender in their language. Kĩĩtharaka does not have derivational morphology that can be separated from number.

The results are also unlikely to be explained by differences in participants' ability to learn or correctly interpret the meanings of gender and number morphemes. Training phase accuracy (in picture-matching and recall tests) for English and Italian speakers were uniformly high. The proportions of correct interpretations of the affixes in the post-experimental survey were also high. Proportions of English and Italian participants reporting the correct interpretations provided for gender morphemes were between 70-75% in animacy-based systems and 85-100% in sex-based systems, and interpretations provided for number morphemes were correct across 85-100% of participants. Most Kĩĩtharaka-speaking participants (93%) also provided correct interpretations for number morphology in the post-experiment questionnaire; however, only 38% of participants provided correct interpretations for the gender morphology, suggesting that the gender markers in this experiment—which, unlike in the other experiments, were attached to stems which were never seen in isolation—were potentially harder to process as separate morphology and they were most likely processed as part of the stem.

It is also worth noting that participants inferred the predicted order regardless of whether the system was prefixal or suffixal, confirming that this preference is not about the sequential order of the morpheme—i.e., which one is linearly first absolutely—but about the which morpheme should be closer to the noun stem. We also found that the preference towards number-closest orders is consistent across sex-based variable gender, and animacy-based fixed gender systems. This suggests that the interpretation of gender across the board leads to a gender-first derivation—i.e., gender combines morphosyntactically with the stem first—separate from number.

Moreover, we generally found that the type of gender system—fixed or variable—did not substantially alter the strength of the preference for gender-closest orders. That said, we did find a slight difference in the likelihood of producing the reversed number-closest order across gender system types: both English and Italian speakers tended to produce this reverse order slightly more often in sex-based variable systems. This might be taken to suggest that gender morphemes are more likely to be treated as separate from the stem when the stem is used to derive related entities (e.g., male vs. female of a given animal as in our sex-based condition). This is potentially in line with theories that posit that gender is derived with the noun stem in fixed systems and with number in variable systems (e.g., De Vincenzi, 1999). However, given the strength of the overall preference, we cannot rule out the possibility that our sample size is insufficient to assert the difference found across fixed and variable gender, and that datapoints driving the effect might be outliers (see fig. 4).

Altogether, these results suggest that speakers' mental representations of gender and number impact ordering preferences. In particular, these representations are such that gender is consistently placed closer to the noun stem than number morphology is. This overwhelming preference cannot be straightforwardly explained by theories which bundle together gender and number features, such as the one proposed for Romance languages in (Ritter, 1993). Nor can it be explained by distributional information alone (Hahn et al., 2022). Rather, this preference mirrors a trend in the typology of separative gender-number systems captured by theories that argue for a representation of gender together with (or closest to) the noun (Alexiadou, 2004; Harris, 1991; Kramer, 2015, 2016a; Mel'čuk, 2013). Of course, the pervasiveness of cumulative gender-number marking in noun affixes still requires explanation. Our findings point toward alternative mechanisms, for example the grammaticalisation of gender systems. We return to this in the general discussion.

Before that, however, we return to the observation that the tendency to place the expression of gender closer to the noun is also found in double plurals (Di Garbo, 2014; Kramer, 2016b). In the examples provided above in section 1.1 for Kinshassa Lingala (see 7), an affix which expresses gender and number cumulatively is placed closer to the stem than a general number affix (which does not express any gender feature). Kramer (2016b) describes a similar tendency for the double plurals found in Amharic (Afro-Asiatic). In Amharic there is a general plural marker which can be used with all stems, (which she calls regular plural) and other plural markers which can only be used only with a restricted set of stems (which she refers to as irregular plurals). Both types of plural markers can co-occur on the same nominal form simultaneously, and in these doubly marked plural nouns, the more restricted number marker is always closer to the stem than the general number marker. This recurrent double plural pattern suggests that number expressions might not all be equivalently represented. In particular, we might expect that general number morphemes might be represented as hierarchically higher than gender-specific (i.e., restricted to a class of stems) number morphemes (Kramer, 2016b) when these are bundled together. Assuming that linear proximity mirrors local structural relationships between morphemes (e.g., Bobaljik, 2012; Embick, 2010), this would explain why they are realised closer to the stem cross-linguistically.

Interestingly, the system Kîîtharaka-speaking participants learned in Study 1 could be somewhat parallel to this. Given their native language experience—i.e., extensive cumulative marking of gender and number—these participants may have learned that the gender affix was a cumulative singular-human class morpheme, and that the number affix was a cumulative plural-inanimate class morpheme. If so, then essentially, the test phase requires them to disregard the class information in the plural morpheme and use it as a general plural at test. This is of course, highly speculative. In the following section we present an experiment explicitly testing ordering preferences of general and gender-specific plural affixes in double plural constructions.

This experiment serves two purposes. First, it allows us to test whether plurals carrying the expression of gender are consistently placed closer to the stem by language learners. If so, this would provide support for the idea that any morpheme carrying gender information should be closer to the stem than any morpheme not carrying such information. This in turn would suggest that distinct representations of number can exist. Second, since both morphemes will express plural (which is more straightforwardly an inflectional category in participants' L1), this experiment will allow us to confirm even more strongly that the results of the first set of experiments are not driven by a bias to place derivational (or more derivational-like) morphology closer to the stem than inflectional morphology.

3 Study 2: ordering biases for general and gender-specific number morphemes

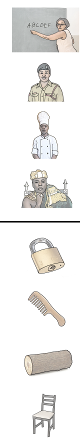
Following our previous study, here we use an artificial language learning paradigm to test whether the inferences learners make about the order of gender-specific (here, gender-encoding) and general (here, non-gender-encoding) number morphology support the idea that these two types of number morphemes are represented distinctly. In particular, gender-specific plurals are hypothesised to be lower in the hierarchy or derived earlier than general plurals, which are higher in the hierarchy (Kramer, 2016b).[22] As in the previous experiments, here we control for any effects of input statistics on participants' ordering preferences. In this case, we explicitly compare order preferences for stems where co-occurrence statistics are balanced, and those which are not. This is explained in more detail below.

3.1 Materials and methods

As in the set of experiments in Study 1 above, here we use an extrapolation design. We train native English speakers on a subset of a nominal system with two gender classes (human and non-human) and two number values (singular and plural). Plurals are overtly marked in this system, and they can be marked with a productive general plural affix or a plural affix that is restricted to a particular gender class. During testing, we ask participants to produce double plurals for the nouns they have been trained on, that is, to use both general and gender-specific plural affixes together.

The experiment features a 2×2 between-subjects factorial design where we manipulate the type of affixation (suffixing or prefixing) and the order of the training blocks of the different plural

[22] As noted above, Kramer (2016b) calls these two types of plurals regular and irregular, however we will use the terms gender-specific and general to avoid confusion since both types can be productive (and are both productive in our experiment).



	SG	PL.SPECIFIC	PL.GENERAL
HUMAN	teek	teek - sa	
	pol	pol-sa	
	chev	chev-sa	chev-gu
	keng	keng-sa	keng-gu
INANIMATE	lak	lak-vi	lak-gu
	cam	cam-vi	cam-gu
	wod	wod-vi	
	sher	sher-vi	

Table 2

Example training artificial lexicon for a suffixal system in Study 2. Each of the two gender-specific plural affixes occurs only with nouns in one gender class; the general plural affix occurs with (a subset of) nouns in both gender classes. Stems that occur with both plural morphemes are referred to as unrestricted stems; stems that occur with only gender-specific plural morphemes are referred to as restricted stems.

markers (general or gender-specific first). Participants were exposed to three blocks of training total, as described in detail below.

3.1.1 Artificial lexicon

Participants were trained on a language with a small lexicon of 8 stems, and 3 affixes. Stems belonged to one of two gender classes, the first containing human referents, the second containing inanimate referents (the same meanings used in the Kîîtharaka experiment in Study 1). There were four stems in each gender class. All three morphemes expressed plural meaning, but one (the general plural) co-occurs with all stems, and two (the gender-specific plurals) co-occur only with stems in a given gender class. All plural affixes are attached to stems, both either as prefixes or suffixes, depending on the condition (randomly assigned for each participant at the beginning of the experiment).

An example of the artificial lexicon participants are trained on is provided in table 2. During training, participants see half of the stems for each class with both general and gender-specific plurals (unrestricted stems) and the other half only with the gender-specific plurals (restricted stems). This allows us to (1) show participants that there is a plural that is not dependent on gender while the other two are, (2) match the frequency of all plural morphemes, and (3) maintain the same point-wise mutual information across affixes and a subset of stems. The latter is discussed further below.

3.1.2 Experimental procedure

The experiment was programmed using *JsPsych* (De Leeuw, 2015) and displayed in participants' browser window. Before starting the experiment proper, participants were told they would learn "how to describe objects and people" in a new language. The training phase was divided into three sub-phases: exposure, picture-matching, and recall. In the final critical testing phase, we ask participants to produce doubly marked plurals and thus add the two affixes (general and gender-specific) to the stems.

During the exposure phase, participants saw images along with orthographically-presented descriptions using a stem alone, or a stem with a single plural affix (either general or gender-specific). In the first block, they were exposed to bare stems, which expressed singular meanings. Participants saw each bare stem twice (i.e., twice for each of the four human stems and twice for each of the four inanimate stems, 16 trials in total, randomly ordered for each participant). In two subsequent blocks they were taught general and gender-specific plurals separately. The order of the general and gender-specific plural training blocks was determined randomly for each participant. Before the first plural training block, participants were told they would be learning"how to form plurals, that is, how to refer to a group of objects or people". Before the second plural training block, participants were told that this is not the only way to refer to groups of objects or people, and that there is another way which they will proceed to learn. In the gender-specific plural block, they saw each of the eight stems four times with the corresponding plural marker (total of 32 trials, randomly ordered for each participant). In the general plural block, they saw only four stems (two assigned to human gender and two to inanimate, see section 3.1.1 above), four times each (total of 16 trials, randomly ordered for each participant).

Importantly, this training regime entails that point-wise mutual information (pmi) between any combination of plural affix (general or gender-specific) and unrestricted stem (those that appear both with both types of plurals) is exactly the same (pmi = 0.585). This is important because it means participants could not use co-occurrence statistics in the input to inform their decision about whether to order gender-specific or general plurals closer to these stem. For restrictive-only stems (which only appear with a gender-specific plural morpheme) the pmi between stems and affixes is higher (pmi = 1.585). Thus in principle participants *could* use this kind of distributional asymmetry to guess the order of plural morphemes. In the test phase (described below) participants were asked to form double plurals with both variable and restrictive-only stems; they do so for unrestricted stems first.

In the next part of the training, the picture matching phase, participants saw two images and a description (using a stem alone or a stem and one plural affix, either general or gender-specific) and had to click the corresponding image. The foil image for each trial was either the wrong gender or the wrong number. Participants saw each stem twice in the singular, twice with the gender-specific plural, and each restricted stem was seen twice in the general plural (40 trials total, 16 singular, 16 gender-specific plural, 8 general plural, randomly ordered for each participant). Co-occurrence statistics were held constant as described for the previous exposure phase.

In the recall phase, participants were tested on the forms they had been trained on. In each trial, they saw an image and had to construct the corresponding form by clicking buttons. The buttons always included the correct stem and two affixes. For plural trials, the two affix buttons were the target affix and a foil, the incorrect gender-specific plural affix. For singular trials, the two affix buttons (both foils) were the general plural affix and the gender-specific plural affix matching the stem. Trials were constructed so that the co-occurrence statistics were maintained as in the previous phases. The general plural was never a foil affix, only a target. Therefore, no explicit evidence is provided in this phase that restricted stems cannot be pluralised with general plurals. Participants saw each of the bare stems or stem+affix combinations they were trained on three times (total of 60 trials, i.e., 24 singular, 24 gender-specific plurals, and 12 general plurals, randomly ordered for each participant).

In the critical testing phase, participants were asked to produce double plurals. Participants were reminded that they "have learned that there are two ways of forming plurals" and were told that "speakers of this language can actually use both together". On each trial, participants saw a plural image and had to construct a doubly-marked form by clicking buttons. The buttons on each trial included the stem, the correct gender-specific plural affix (i.e., the one that matches the gender class of the stem) and the general plural affix. Participants had to use all three buttons and could not submit their answer until they did. No other feedback was provided. In the first block of testing, participants only saw unrestricted stems, that is, stems which appeared with both types of plurals during training. They were asked to produce double plurals three times for each unrestricted stem (total of 12 trials, randomly ordered for each participant). Recall that for these items, the frequency of all single affix-stem combinations during training is the same (pmi = 0.585) and there is thus no evidence that one or the other plural affix is more closely associated with a stem based on co-occurrence statistics. In the second block of testing, participants saw restricted stems. They were asked to produce double plurals three times for each restricted stem (total of 12 trials, randomly ordered for each participant). For these items, there is a higher point-wise mutual information between the gender-specific plural affixes and stems (pmi = 1.585) compared to general plurals and stems. There was no break between these blocks.

3.1.3 Participants

English-speaking participants were recruited through Prolific as in our previous experiments and paid 10 GBP/h for participation in the study, which took approximately 15 minutes. We used Prolific screening criteria to include only native English-speaking participants who were raised monolingual and whose primary language is still their native language. As per our preregistration, participants whose accuracy was lower than 90% in the last block of the single morpheme production phase during training were excluded. We also excluded participants who did not produce both affixes (as both suffixes or prefixes) during testing. After exclusions (N =35), a total of 120 participants' data was analysed (30 in each condition of our 2x2 factorial design).

3.1.4 Data analysis

We ran Bayesian logistic regression models to test whether participants were more likely to produce gender-specific plural morphemes (containing gender information) closer to the stems than the general plural morpheme (without gender information). The model fixed effects were affix position (treatment coding, suffixal vs prefixal), training block order (dummy coded, irregular first = 0, and regular first = 1), and stem type (treatment coded, with restricted stems vs. unrestricted stems), and their interactions. As random effects, we included intercepts for participant and gender class (human or non-human), and by-participant slopes for the effects of stem type.

3.1.5 Transparency and openness

All data and analyses reported are available at osf.io/genderNumber. The preregistered hypothesis, design and analysis plan can be found in osf.io/doublePlural.[23]

[23]Note that this study was preregistered as a replication of an earlier (unpublished) study. We ran this replication with modifications to the sample size and the design method because the original study was not designed to have enough power to detect any differences between training orders (gender-specific first or general first). The preregistration for the

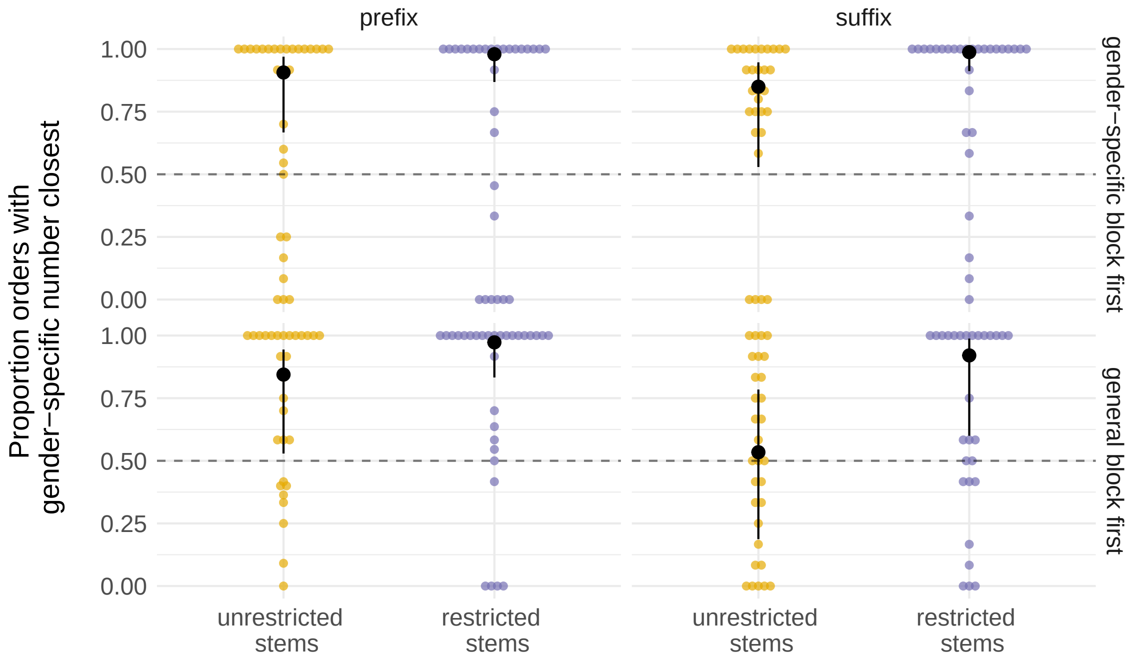


Figure 5

Average proportion of responses with gender-specific plural ordered closest to the stem among English-speaking participants in Study 2, by affix type (suffix or prefix), training block order (gender-specific or general plural first) and stem type (unrestricted or restricted). Coloured dots represent participants' individual scores; black dots represent the model's predicted mean accuracy scores and the error bars represent the model's predicted 90% credible intervals. Dashed lines represent the chance level.

3.2 Results

Based on our preregistered hypothesis, we predicted that (1) participants would prefer to place gender-specific plurals closer to the stem than general plurals, that (2) they would do so independently of the affix position and the input co-occurrence statistics, but (3) that this preference would potentially be stronger with restricted stems than with unrestricted stems as the dependency between gender-specific plurals and restricted stems was stronger.

A visual inspection of the results, shown in Figure 5, suggests that our predictions are generally met. Overall, participants show a preference to place gender-specific plurals closer to the stem than general plurals, and this effect is stronger for those stems which appeared with only a gender-specific plural during training (i.e., restricted stems). However, results also suggest that this might interact with affix position—the difference between restricted and unrestricted stems is greater for suffixal conditions. The model results confirm these observations. The intercept estimates provide strong evidence in favour of a preference for ordering the gender-specific plural closer to unrestricted stems in the prefixed condition when gender-specific plurals are taught first

original study can be found at osf.io/doublePluralExcl, and its data analysis can also be found in the analysis script within the supplementary materials in osf.io/genderNumber.

($\hat{\beta} = 2.203$, 90%CI = $[0.695, 3.456]$, $SE = 0.875$, $P(\hat{\beta} > 0) = 0.977$). We did not find robust evidence for an effect of training order (general first: $\hat{\beta} = -0.593$, 90%CI = $[-1.823, 0.582]$, $SE = 0.735$, $P(\hat{\beta} < 0) = 0.791$; general first and suffix: $\hat{\beta} = -1.003$, 90%CI = $[-2.556, 0.501]$, $SE = 0.934$, $P(\hat{\beta} < 0) = 0.863$) or affix type ($\hat{\beta} = -0.552$, 90%CI = $[-1.776, 0.637]$, $P(\hat{\beta} < 0) = 0.775$); this suggest that the preference for placing gender-specific plurals closer to unrestricted stems applies comparably across conditions.

The model results further suggest that the preference for placing gender-specific plurals closer to stems increases for restricted stems as predicted ($\hat{\beta} = 1.637$, 90%CI = $[0.479, 2.929]$, $SE = 0.745$, $P(\hat{\beta} > 0) = 2.451$). This effect is stronger in suffix conditions ($\hat{\beta} = 1.077$, 90%CI = $[-0.262, 2.451]$, $SE = 0.829$, $P(\hat{\beta} > 0) = 0.907$), and potentially driven by the greater difference found in the suffix condition when the general plural is taught first. This is also shown in Figure 5: the difference between restricted and unrestricted stems is greatest in the condition with suffixes and general plural training first. Here, participants have a strong preference for gender-specific number closest for restricted stems but they have no preference for either order with unrestricted stems. We nevertheless do not find strong support for these observations in our model, only a mild tendency (general first and suffix: $\hat{\beta} = -1.003$, 90%CI = $[-2.556, 0.501]$, $SE = 0.934$, $P(\hat{\beta} < 0) = 0.863$; general first, suffix and restricted stems: $\hat{\beta} = -0.648$, 90%CI = $[-2.293, 0.954]$, $P(\hat{\beta} < 0) = 0.746$). We will discuss possible explanations for this asymmetry across conditions below (see 4.2).

3.3 Discussion

In Study 2, we explored whether participants' preference for ordering gender closer to the stem than number would extend to a system in which the morpheme encoding gender information cumulatively expressed number. In this system, plural meaning could be expressed using either marker alone or both markers together. Participants tended to place gender-specific number markers (which appeared only with the nouns assigned to a specific gender class) closer to the stem than general number markers (which could appear with stems of either class). We also found some evidence for an influence of co-occurrence statistics: the preference for gender-closest order was greater for restricted stems.

However, it is also worth noting that, despite being explicitly told that morphemes expressed plural meaning, in our post-experiment questionnaire, participants varied in whether they explicitly interpreted both morphemes as having a plural meaning: 64% of participants explicitly identified plural as the meaning of the general plural affix. This was somewhat lower than in our previous experiments, and probably reflects the unusual nature of this redundant plural system for English speakers. Around 58% identified plural as (part of) the meaning of (both) gender-specific plural affixes, with many instead identifying only the animacy meanings (i.e., they reported that the gender-specific plural affixes meant "person" or "object")—while 53% interpreted animacy as part of the meaning of the two gender-specific markers, only 36% of participants identified both the animacy and plural meanings together in the same markers.[24]

Of course, descriptions (or failures to correctly describe) an unfamiliar system are not necessarily meaningful. That is, they do not necessarily tell us that participants failed to represent the

[24] We ran a (non-preregistered) regression model to test for differences in the accuracy of the gender and plural descriptions for the different types of affixes (gender in gender-specific, plural in gender-specific and plural in general plural affixes) across conditions. We found that the interpretation accuracy was comparable across meanings and conditions. Further details of the model and results can be found in the analysis script at osf.io/genderNumber/.

meanings as plurals. But, it suggests the possibility that at least some participants may have reinterpreted the gender-specific plural as a gender marker at test. If participants are reinterpreting the gender-specific plural as a gender marker, then these results simply replicate those of Study 1. For those participants who do in fact retain the intended meaning of the markers (perhaps despite their problem describing the system), then our results replicate the preference for gender-closest orders found in the first study but in a double plural system.[25] This supports the idea that the representation of number may depend on whether it is cumulative with gender or not. Importantly, these results also suggest that the ordering preferences we found in Study 1 were not driven by a preference for ordering derivational morphemes closer to the stem than inflectional morphemes. We discuss both these points in more detail below.

4 General Discussion

In this paper, we started with the observation that although morphological systems vary along many dimensions, there are nevertheless a number of apparent regularities. In the domain of grammatical gender, for example, there are common semantic bases of classification, and regularities in the interaction with number morphology. Specifically, gender and number are often realised via cumulative exponence, and in the relatively rare cases where they are not, morphemes encoding gender are closer to the noun stem than morphemes encoding number. These regularities have influenced theories of the morphosyntactic representation of gender. For example, split models argue that gender is represented more locally to the noun stem—either within the nominal or nominalising head itself (e.g., Alexiadou, 2004; Harris, 1991; Kihm et al., 2005; Kramer, 2015), or in a distinct projection hierarchically dominating the nominal head](e.g., Antón-Méndez et al., 2002; Carminati, 2005; Panagiotidis, 2018; Picallo, 1991). Assuming that linear proximity mirrors local dependencies between morphemes (e.g., Bobaljik, 2012; Embick, 2010), these models predict that when morphemes are linearised, gender will be closest to the noun stem. By contrast, what we have called bundling models argue that gender does not project separately in the morphosyntax from number (e.g., Carstens, 2003; Ritter, 1993). These models don't make any predictions about the relative order of gender and number (see however, section 4.2), rather they more naturally capture the trend for cumulative exponence. In stark contrast with these models, alternative accounts of morphological ordering and fusion are based not on the nature of hierarchical representations per se, but on distributional information (Hahn et al., 2021; Hay & Plag, 2004; Rathi et al., 2022; Ryan, 2010). Under these accounts, for example, co-occurrence frequencies between morphemes would explain the kinds of regularities noted above.

Across two sets of experiments, here we used artificial language learning experiments to test whether learners trained on a novel system of gender and number would exhibit a preference to order gender-encoding morphemes closer than (exclusively) number-encoding morphemes. Crucially, our experiments were designed so that participants had either no, or minimal information in the input that they could use to determine order. Thus generally, we were interested in whether a mental representation of gender as more local to the noun than number would determine the order they used. In Study 1, we tested English, Italian, and Kĩĩtharaka speakers on a system with two genders and two numbers. For each feature, one was marked overtly and the other was not, and we held

[25] A model including only the data from those participants who described the intended meaning still shows that participants prefer gender-specific closest orders ($P(\hat{\beta} > 0) = 0.986$) and more so with restricted stems $P(\hat{\beta} > 0) = 0.982$. See analysis script in osf.io/genderNumber/.

out meanings where both overt morphemes co-occurred. For English- and Italian-speaking participants, we designed systems which balanced the frequency of co-occurrence of all stem+morpheme combinations. For Kîîtharaka-speaking participants, we designed a system where stems were always marked for their particular gender (more similar to natural languages). We found a strong preference to order gender closer to the stem than number in all cases—across all three groups, for both prefixal and suffixal morphemes, for both sex-based (variable) and animacy-based (fixed gender) systems. These results therefore provide strong support that, independent of distributional information, when they are separative, gender is ordered closer to the stem than number.

In Study 2, we explored a somewhat different system, inspired by a natural language phenomenon sometimes called double plurals. In these systems, there is a general plural marker than can occur with any stem, and another that is restricted—for example, to stems with a particular gender. We created an artificial language modelling this kind of system, and found a preference in English-speaking learners to place gender-specific number markers (i.e., number markers that encode gender) closer to the stem than general number markers (i.e., markers than encode exclusively number). These results thus replicate the the preference for gender-closest orders found in the first study but in a double plural system.

4.1 The derivation/inflection divide

Before we turn to the implications of our results for theories of gender specifically, and morpheme order more generally, we first discuss a potential question about what drives participants' preferences in our experiments. In particular, one potential concern in interpreting the overwhelming preference for gender-closest orders in Study 1 is whether participants interpreted gender morphology as derivational and number morphology as inflectional. If this were the case, the observed preference could reflect the more general cross-linguistic tendency to place derivational morphology closer to the root than inflectional morphology. This is problematic since English and Italian both follow this tendency, *and* they both have derivational morphemes that can be used to express gender (e.g., English 'lion-s/lion-ess-es' and Italian 'leon-i/leon-ess-e'). Participants' preferences could thus simply reflect L1 knowledge. Even aside from this, distinguishing between derivation and inflectional is difficult (e.g., Dressler & Doleschal, 1990).

However, several aspects of our findings suggest that the derivational flavour of gender (either in general, or for our participants) cannot explain our results. First, in Study 1 we observed the same strong preference for gender-closest orders across all populations we tested. This includes Kîîtharaka speakers, who have no sex-based derivational gender, and no evidence for the relative order of gender and number separately through derivation and inflection in the nominal domain. Further, if anything, we found that the preference for gender-closest orders was slightly weaker in sex-based (variable) compared to animacy-based (fixed) systems. If participants were processing gender as derivational and number as inflectional, we would expect to see the opposite. Finally, the results from Study 2 show that even when gender information is encoded in a clearly inflectional number morpheme, the preference persists. Even if some participants did reanalyse the gender-specific number marker as a gender marker at test, it is unlikely that this morpheme would be interpreted as derivational given that the referent of the stem alone and the stem with the (reanalysed) gender marker are exactly the same. Setting this issue aside then, we turn to the implications of our results on theories of the morphosyntactic representation of gender.

4.2 Implications for theories of gender

The results of Study 1 suggest that mental representations of gender and number may impact morpheme ordering preferences. Specifically, our results support theories in which gender is represented more locally to the nominal stem than number. These results are in line with previous work using artificial language learning methods to provide evidence linking mental representations of syntax and morphosyntax to preferences for particular linearisation patterns (Herce, Saldana, Mansfield, & Bickel, 2023; Maldonado et al., 2020; Mansfield et al., 2022; Martin, Adger, Abels, Kanampiu, & Culbertson, 2024; Martin, Holtz, Abels, Adger, & Culbertson, 2020; Saldana, Kirby, Truswell, & Smith, 2019; Saldana et al., 2021). In terms of the specific theories of gender discussed above, our results most naturally align with theories adhering to a split model where gender projects separately from number (e.g., Alexiadou, 2004; Harris, 1991; Kramer, 2015; Panagiotidis, 2018).[26] Our results are less straightforwardly captured by theories which bundle together gender and number features in the morphosyntax (e.g., Carstens, 2003; Ritter, 1993). If gender and number are represented together, we would expect to observe more variation in the ordering of separative gender and number morphology in our experiments—and potentially no preference at all. In order to account for the clear preference we in fact found, bundling models would require some ad-hoc ordering mechanism forcing number to linearise outside of gender despite being bundled together (for an example of this, see Ritter, 1993).

In addition to testing the different predictions of split and bundling models of gender, we also explored potential differences between fixed and variable gender systems. Specifically, we compared an animacy-based system in which stems were shared idiosyncratically across classes, or not shared at all, to a gender-based system in which stems expressed a consistent meaning (female vs. male) across classes. Several formal theories assume that variable gender projects higher than fixed gender, either together with number (e.g., Carminati, 2005; De Vincenzi, 1999) or separate from number (e.g., Panagiotidis, 2018). In these theories, fixed gender is considered to be part of the nominal(ising) head or part of the lexical entry, in contrast to variable gender.

In Study 1, we found a slightly higher preference—although still very low—for alternative number-closest orders in sex-based systems. In our preregistered initial statistical models, which separately analysed data from the experiments with the English and Italian populations, we found stronger evidence for this preference for our Italian-speaking population. However, when we analysed the combined results, we found moderate support for a lower preference of gender-closest orders in sex-based systems across both linguistic populations. While these results are in line with the aforementioned predictions, the overwhelming preference for gender-closest orders suggests that if there is indeed a difference, it is quite weak.

Turning to Study 2, participants' preference for placing gender-specific plural markers closer to the stem than general plural markers is consistent with the idea that number morphol-

[26] Our experiments were not explicitly designed to distinguish between theories in which gender is located within the nominalising head (e.g. Kramer, 2015), and those which treat gender as heading its own projection outside the nominal (e.g Picallo, 1991). That said, one could argue that participants in Study 1 are unlikely to have treat separative gender morphology as part of the lexical entry of the stem, particularly when stems were shared across classes (in experiments with English and Italian speakers), and were semantically interpretable (in the sex-based, variable system). As discussed below, the preference for gender-closest order was very similar in these conditions as when the stems were not shared (in our experiment with Kĩĩtharaka speakers), or where not semantically interpretable (in the animacy-based, fixed system). Thus our results might be taken to suggest at the very least that even if both types of representations are possible, there are unlikely to lead to differences in locality between gender and the stem.

ogy can potentially occupy two different loci. As argued by Kramer (2016b), for example, number morphology which encodes gender information may be locally (and therefore linearly) closer to the stem than pure number morphology. This mirrors the tendency in natural language double plural systems such as those found in Kinshasa Lingala or Amharic (Di Garbo, 2014; Kramer, 2016b). In light of this, the results of Study 1 can be seen as specific to the highest projection of number (e.g., NumP). The results of Study 2 would potentially be compatible with the idea that bundling of gender and number might also occur in the morphosyntax, but could represent a distinct, and lower, projection.

4.3 Cognitive and distributional biases in morpheme ordering

Above we have argued that participants' robust preference for placing gender closer to the stem than number reflect their mental representations of gender and number. Specifically, we have argued that the *meaning and function* of gender and number lead to different dependency relations with the stem, represented in a morphosyntactic hierarchically (Bybee, 1985; Embick, 2010; Maldonado et al., 2020; Saldana et al., 2021, see also, e.g.,). This finding is in line with previous research, especially Saldana et al. (2021) and Culbertson et al. (2020), who highlight the role of meaning in driving ordering preferences for number and case morphology, and nominal modifiers respectively. Here we have focused on very specific theoretical accounts that are in principle independent of these previous findings. However, it is worth emphasising that we are not endorsing a particular view of the origin of these representations. For example, they could reflect innate linguistic representations, or they could reflect learned representations, e.g., about how concepts or meanings are constructed (e.g., Culbertson et al., 2020).

In addition to these representations, whatever their ultimate origin, we also found some evidence that morpheme order preferences are impacted by dependency relations based on co-occurrence statistics (Hahn et al., 2022; Hay & Plag, 2004; Ryan, 2010). In Study 1, we have evidence that meaning and function drive sufficiently strong preferences that adding evidence from co-occurrence statistic does not make a difference. In particular, in our experiment with Kîîtharaka speakers, stems were not shared across classes, and thus each stem always occurred marked with its gender (in the marked gender class). By contrast, in the experiments with Italian and English speakers, stems were shared, co-occurrence statistics were balanced. Nevertheless, the preference for gender-closest order was equally strong across all three experiments.However, in Study 2, there is more variability in participant's response—i.e., the preference for gender-closest order was not at ceiling. Here, we had a subset of stems for which the co-occurrence statistics between morphemes was balanced (unrestricted stems), and a subset where mutual information between the stem and the gender-specific plural marker was higher (restricted stems). This stronger statistical dependency effected participants' behaviour: it led to a stronger preference for placing the gender-specific plural marker closer to the stem than the general plural marker (see fig. 5). These results suggest that participants' representation of dependency relations between morphemes can be affected both by their meaning and function, and evidence from co-occurrence statistics (see also Saldana et al., 2021).[27]

[27] It is worth noting that in Study 2, the testing block with unrestricted stems always preceded the block with restricted stems. Our results might reflect this to some extent: participant might become less variable over time, and participants' weak bias towards placing gender information closer to the stem might have become more categorical over time (i.e., eventually they chose their preferred form each time). If anything, this would have weakened the effect of co-occurrence statistics.

4.4 Returning to the typology

A major aim of the studies reported here was to link individual-level cognitive representations and mechanisms to cross-linguistic tendencies in morpheme order (following previous recent research, e.g., Hahn et al., 2022; Maldonado et al., 2020; Mansfield et al., 2022; Saldana et al., 2021). In this particular case, we focused on the tendency for gender to surface closer to the stem than number when a language has separative morphology for these two categories (e.g., Fuchs et al., 2015). However, we also noted that separative or partially separative expressions of gender and number are very rare (see 2). Even within these rare cases, there are examples of languages which allow both gender-closest *and* number closest orders. Similarly, while we have not come across a description of double plural system in which a general plural is placed closer to the stem than a gender-restricted plural, the evidence is again very scarce (Di Garbo, 2014; Kramer, 2016b). Thus the tendency for separative morphology to conform to gender-closest orders is not absolute and the typological evidence alone does not allow us to robustly assess the strength of this tendency. For these reasons, experimental methods like artificial language learning can offer a crucial source of additional evidence. These experiments complement natural language data and allow us to make claims about the causal link between human cognition and typology. Biases in representations, like the ones we have uncovered here have the potential to shape languages at each generation, shaping how languages change over time (Bickel, 2015; Blythe & Croft, 2021; Kirby et al., 2015; Reali & Griffiths, 2009; Smith, 2018; Thompson et al., 2016).

Of course, the strong preference that we observe in language learners to place gender-information closer to the stem than number leaves the pervasiveness of cumulative gender-number marking in natural languages unexplained. One possibility is that tendency for gender and number to be cumulative can be better accounted for by the way in which grammatical gender systems are likely to emerge.

4.4.1 The emergence of grammatical gender systems

The emergence of grammatical gender systems is a long studied phenomenon and the literature is too vast to discuss it in detail here (Aikhenvald, 2016; Bopp, 1816; Corbett, 1991; Di Garbo, 2014; Di Garbo & Miestamo, 2019; Fodor, 1959; Greenberg, 1978; Luraghi, 2011; Matasović, 2004; Nichols, 1992; Royen, 1929; Seifart, 2005). For the purposes of this brief discussion, it suffices to say that there are two main scenarios. In one scenario, gender systems originate from classifiers and classificatory nouns that grammaticalise into agreement markers and, eventually, form the basis of gender markers on nouns as well (e.g., Corbett, 1991; Greenberg, 1978). In an alternative scenario the development of agreement markers precedes the development of any classificatory distinctions (Nichols, 1992).

Greenberg (1978) proposed that gender agreement markers develop from demonstratives and then articles (see also Corbett, 1991). Crucially, the demonstratives must either have marked gender distinctions or their distribution should at least be based on a (covert) gender distinction (e.g., having its use restricted to a particular set of nouns). The articles that develop from these demonstratives (and potentially from succeeding anaphoric pronouns; Corbett, 1991; Greenberg, 1978) begin as definite articles, but later expand to include non-definite determination, and finally become noun markers without any connection to specificity or definiteness. Within his theory, however, the development of such classifying demonstratives in the first place still requires explanation. Greenberg (1978) proposes that classifying demonstratives could develop from classifiers, and from numeral classifiers in particular. In a similar proposal Corbett (1991) highlights the role of classifiers, arguing that the ultimate source of gender systems is nouns with classificatory possibilities (see also Di Garbo & Miestamo, 2019; Heine & Vossen, 1983) which later develop into classifiers before potentially giving rise to demonstratives.

These explanations of the rise of gender thus depend on a transition period between a classifier systems and a gender system. However, Nichols (1992) notes that classifiers and gender do not tend to overlap (for further supporting evidence with larger lingusitic coverage, see Grambank, Skirgård et al., 2023, in particular, features GB038, GB057, GB051-53, GB057-58, GB172, GB321). Nichols (1992) instead proposes an alternative scenario whereby classificatory distinctions infiltrate preexisting person and/or number agreement patterns whose distribution might be based on morphologically covert (i.e., not overtly realised) distinctions according to "highly cognitively salient" domains such as animacy or perceived biological sex in animate entities. In this scenario, there are three prerequisites for grammatical gender to arise: a covert system of gender distinctions, and a potentially recruitable formal distinction and preexisting agreement patterns. Thus contrary to the proposals of Greenberg (1978) and Corbett (1991), in the mechanism proposed by Nichols (1992), agreement triggers noun classification (rather than vice versa).

Whichever the scenario, all accounts assume that in its origin, a gender system encroaches on lexical or morphological material where number distinctions are already present, and the succeeding forms that evolve into gender marking will most likely also carry the number features along with them. In the classification-before-agreement scenario (Corbett, 1991; Greenberg, 1978; Heine & Vossen, 1983), gender develops from classifiers and determiners, either of which would already have distinctions based on number. In the agreement-before-classification scenario (Nichols, 1992), a system of number agreement—together or not with a system of person agreement—is a prerequisite and classification will be built on the available formal marking in such an agreement system. Across these scenarios, cumulative marking is thus in some sense the default. The most likely route for gender and number to become (or at least be able to be described as) separative is if an initial cumulative gender-number marker can be reanalysed as separative once gender and number agreement systems are in place (as is the case in the later development of already stable gender systems in some Romance languages, e.g., from Latin to Spanish). Alternatively, as observed in Kinshasa Lingala (see 7; Bokamba, 1977; Di Garbo, 2014), a plural class marker can be reanalysed as a general plural, which is placed together with other plural class markers resulting in redundant number marking. In this case, the general plural has lost its gender function and can thus be described as a separative number marker. In the future, plural class markers which frequently co-occur with the general plural could be reanalysed as separative gender morphology—just as some of the participants in our Study 2 appear to have done.

5 Conclusion

In this paper, we tested a hypothesis based on cross-linguistic data and theories of morphosyntax predicting that morphemes encoding gender should be placed closer to the stem than morphemes encoding number. In a series of artificial language learning experiments, we found clear evidence supporting this prediction, across a number of different types of systems, regardless of whether morphemes were prefixal or suffixal, and with equal strength independent of participants' native language experience. We interpret this result as providing support for theories of gender which treat this morphosyntactic feature as representationally distinct from number, and locally (hierarchically) closer to the nominal stem. Importantly, the ordering preference we found was based on the meaning and function of gender *not* distributional information about the frequency of co-occurrence in the languages participants were trained on. However, we also found some evidence that when learners' preferences were weaker—when gender information was itself encoded by a redundant plural marker—co-occurrence statistic did influence order. These findings shed light on the general mechanisms driving ordering preferences, and add to the growing body of research using artificial language learning experiments to link human cognition to language structure.

Acknowledgements

The work has been partially funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 757643, held by JC) and the NCCR Evolving Language, Swiss National Science Foundation (Agreement Nr. 51NF40_180888). CS is currently supported by the Beatriu de Pinós programme funded by the Generalitat de Catalunya (agreement No. 2022 BP 00101) and the European Union's Horizon 2020 research and innovation programme under the MSCA grant agreement No 801370.

Ethics

The studies were approved the Ethics Committee of the School of Philosophy at the University of Zurich (Ref No 22.10.12) and the PPLS Research Ethics Committee at the University of Edinburgh (Ref No 423-2122/2).

References

Aikhenvald, A. Y. (2016). *How gender shapes the world.* Oxford University Press.

Alderete, J., Jebbour, A., Kachoub, B., & Wilbee, H. (2015). Tashlhiyt berber grammar synopsis. *Simon Fraser University*.

Alexiadou, A. (2004). Inflection class, gender and DP internal structure. *Explorations in nominal inflection*, 21–50. doi: 10.1515/9783110197501.21

Antón-Méndez, I., Nicol, J. L., & Garrett, M. F. (2002). The relation between gender and number agreement processing. *Syntax*, *5*(1), 1–25. doi: 10.1111/1467-9612.00045

Baerman, M., Brown, D., & Corbett, G. G. (2005). *The syntax-morphology interface: A study of syncretism* (Vol. 109). Cambridge: Cambridge University Press.

Bickel, B. (2015). Distributional typology: Statistical inquiries into the dynamics of linguistic diversity. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Linguistic analysis, 2nd edition* (pp. 901–923). Oxford University Press. doi: 10.1093/oxfordhb/9780199677078.013.0046

Bickel, B., Nichols, J., Zakharko, T., Witzlack-Makarevich, A., Hildebrandt, K., Rießler, M., . . . Lowe, J. B. (2023). *The autotyp database (v1.1.0).* Retrieved from `https://doi.org/10.5281/zenodo.6793367` doi: 10.5281/zenodo.6793367

Blythe, R. A., & Croft, W. (2021). How individuals change language. *Plos one*, *16*(6), e0252582. doi: 10.1371/journal.pone.0252582

Bobaljik, J. D. (2012). *Universals in comparative morphology: Suppletion, superlatives, and the structure of words* (Vol. 50). MIT Press.

Bokamba, E. G. (1977). The impact of multilingualism on language structures: the case of central africa. *Anthropological Linguistics*, *19*(5), 181–202.

Bopp, F. (1816). *Über das konjugationssystem der sanskritsprache in vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen sprache*.

Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form.* Philadelphia, PA: John Benjamins.

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. doi: 10.32614/RJ-2018-017

Carminati, M. N. (2005). Processing reflexes of the feature hierarchy (person> number> gender) and implications for linguistic theory. *Lingua*, *115*(3), 259–285. doi: 10.1016/j.lingua.2003.10.006

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, *76*(1). doi: 10.18637/jss.v076.i01

Carstens, V. (2003). Rethinking complementizer agreement: Agree with a case-checked goal. *Linguistic inquiry*, *34*(3), 393–412. doi: 10.1162/002438903322247533

Carstens, V., et al. (2010). Implications of grammatical gender for the theory of uninterpretable features. *Exploring crash-proof grammars*, *3*, 31.

Corbett, G. G. (1991). *Gender*. Cambridge University Press.

Corbett, G. G. (2000). *Number*. Cambridge University Press.

Corbett, G. G. (2013). Sex-based and non-sex-based gender systems (v2020.3) [Data set]. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.7385533` doi: 10.5281/zenodo.7385533

Crisma, P., Marten, L., Sybesma, R., et al. (2011). The point of bantu, chinese and romance nominal classification. *Rivista di Linguistica*, *23*(2), 251–299.

Culbertson, J., Schouwstra, M., & Kirby, S. (2020). From the world to word order: Deriving biases in noun phrase order from statistical properties of the world. *Language*, *96*(3). doi: 10.1353/lan.2020.0045

De Belder, M., & Van Koppen, M. (2016). One module, different levels of merge: An (n) compounds in dutch. *Studia Linguistica*, *70*(1), 1–33.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods*, *47*(1), 1–12. doi: 10.3758/s13428-014-0458-y

De Vincenzi, M. (1999). Differences between the morphology of gender and number: Evidence from establishing coreferences. *Journal of Psycholinguistic Research*, *28*, 537–553. doi: 10.1023/A:1023272511427

Di Garbo, F. (2014). *Gender and its interaction with number and evaluative morphology: An intra-and intergenealogical typological survey of africa*. Unpublished doctoral dissertation, Department of Linguistics, Stockholm University.

Di Garbo, F., & Miestamo, M. (2019). The evolving complexity of gender agreement systems. In F. D. Garbo, B. Olsson, & B. Wälchli (Eds.), *Grammatical gender and linguistic complexity.* Language Science Press.

Dressler, W. U., & Doleschal, U. (1990). Gender agreement via derivational morphology. *Acta linguistica hungarica*, *40*(1/2), 115–137.

Embick, D. (2010). *Localism versus globalism in morphology and phonology*. MIT Press.

Faust, N., & Lahrouchi, M. (2022). The locus of gender in tashlhiyt berber nouns. In *Workshop on prefixes vs. suffixes in afroasiatic.*

Fodor, I. (1959). The origin of grammatical gender. *Lingua*, *8*, 186–214.

Fuchs, Z., Polinsky, M., & Scontras, G. (2015). The differential representation of number and gender in spanish. *The linguistic review*, *32*(4), 703–737.

Futrell, R., Dyer, W., & Scontras, G. (2020). What determines the order of adjectives in english? comparing efficiency-based theories using dependency treebanks. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 2003–2012).

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336–10341.

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological science*, *24*(7), 1079–88. doi: 10.1177/09567976124637

Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of language* (p. 73-113). Cambridge, MA: MIT Press.

Greenberg, J. (1978). How does a language acquire gender markers. In C. F. Joseph Greenberg & E. Moravcsik (Eds.), *Universals of human language* (Vol. 3, pp. 47–82). Standford University Press.

Hahn, M., Degen, J., & Futrell, R. (2021). Modeling word and morpheme order in natural language as an efficient tradeoff of memory and surprisal. *Psychological Review*, *128*(4). doi: 10.1037/rev0000269

Hahn, M., Jurafsky, D., & Futrell, R. (2020). Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences*, *117*(5), 2347–2353.

Hahn, M., Mathew, R., & Degen, J. (2022). Morpheme ordering across languages reflects optimization for processing efficiency. *Open Mind*, *5*, 208–232. doi: 10.1162/opmi_a_00051

Harris, J. W. (1991). The exponence of gender in Spanish. *Linguistic inquiry*, *22*(1), 27–62. Retrieved from `https://www.jstor.org/stable/4178707`

Hawkins, J. A., & Cutler, A. (1988). Psycholinguistic factors in morphological asymmetry. In J. A. Hawkins (Ed.), *Explaining language universals* (pp. 280–317). Oxford: Blackwell.

Hay, J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics*, *39*(6), 1041–1070. doi: 10.1515/ling.2001.041

Hay, J., & Plag, I. (2004). What constrains possible suffix combinations? on the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language & Linguistic Theory*, *22*(3), 565–596. doi: 10.1023/B:NALA.0000027679.63308.89

Heine, B., & Vossen, R. (1983). On the origin of gender in eastern nilotic. In *Nilotic studies: Proceedings of the international symposium on languages and history of the nilotic peoples, cologne* (pp. 245–268).

Herce, B., Saldana, C., Mansfield, J., & Bickel, B. (2023). Naturalness is gradient in morphological paradigms: Evidence from positional splits. *Glossa*, *8*(1), online. doi: 10.16995/glossa.9280

Jenkins, S., & Jenkins, J. (2000). *A Grammar of W̃eỹ* (Tech. Rep.). SIL Senegal.

Kanampiu, P., Martin, A., & Culbertson, J. (under review). Semantic and morphophonological productivity of kîîtharaka gender system: A quantitative study.

Kihm, A., Cinque, G., & Kayne, R. S. (2005). Noun class, gender, and the lexicon-syntax-morphology interfaces. In G. Cinque & R. S. Kayne (Eds.), *The oxford handbook of comparative syntax* (Vol. 40, pp. 459–512). Oxford, Oxford University Press.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Koopman, H. (2003). The locality of agreement and the structure of the dp in maasai. In W. E. Griffin (Ed.), *The role of agreement in natural language: Tls 5 proceedings* (p. 207—227). University of Texas, Department of Linguistics.

Kramer, R. (2015). *The morphosyntax of gender* (Vol. 58). Oxford University Press.

Kramer, R. (2016a). The location of gender features in the syntax. *Language and linguistics Compass*, *10*(11), 661–677. doi: 10.1111/lnc3.12226

Kramer, R. (2016b). A split analysis of plurality: Number in amharic. *Linguistic Inquiry*, *47*(3), 527–559.

Lowenstamm, J. (2008). On little n, $\sqrt{}$, and types of nouns. In J. Hartmann, V. Hegedus, & H. van Riemsdijk (Eds.), *Sounds of silence: Empty elements in syntax and phonology* (pp. 105–144). Elsevier Amsterdam.

Luraghi, S. (2011). The origin of the proto-indo-european gender system: Typological considerations. *Folia Linguistica*.

Maldonado, M., Saldana, C., & Culbertson, J. (2020). Learning biases in person-number linearization. In *Proceedings of the 50th Annual Meeting of the North East Linguistic Society* (Vol. 2, pp. 163–176). Amherst, MA: University of Massachusetts GLSA. (Preprint doi: 10.31234/osf.io/5s2r8)

Mansfield, J., Saldana, C., Hurst, P., Nordlinger, R., Stoll, S., Bickel, B., & Perfors, A. (2022). Category clustering and morphological learning. *Cognitive Science*, *46*(2), e13107. doi: 10.1111/cogs.13107

Martin, A., Adger, D., Abels, K., Kanampiu, P., & Culbertson, J. (2024). A universal cognitive bias in word order: Evidence from speakers whose language goes against it. *Psychological Science*. doi: 10.1177/09567976231222

Martin, A., Holtz, A., Abels, K., Adger, D., & Culbertson, J. (2020). Experimental evidence for the influence of structure and meaning on linear order in the noun phrase. *Glossa*, *5*(1), 1-21.

Matasović, R. (2004). *Gender in indo-european*. Winter. Retrieved from https://books.google.es/books?id=jKEcAQAAIAAJ

Mel'čuk, I. A. (2013). The notion of inflection and the expression of nominal gender in Spanish. *Studies in Language. International Journal sponsored by the Foundation"Foundations of Language"*, *37*(4), 736–763. doi: 10.1075/sl.37.4.02mel

Nichols, J. (1992). *Linguistic diversity in space and time*. University of Chicago Press.

Panagiotidis, P. (2018). (grammatical) gender troubles and the gender of pronouns. In G. Z. Éric Mathieu Myriam Dali (Ed.), *Gender and noun classification.* Oxford University Press.

Picallo, M. C. (1991). Nominals and nominalizations in catalan. *Probus*, *3*(3), 279–316. doi: 10.1515/prbs.1991.3.3.279

Picallo, M. C. (2008). Gender and number in romance. *Lingue e linguaggio, Rivista semestrale*, 47–66. doi: 10.1418/26709

Rathi, N., Hahn, M., & Futrell, R. (2022). Explaining patterns of fusion in morphological paradigms using the memory–surprisal tradeoff. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44).

Reali, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, *111*(3), 317–328. doi: 10.1016/j.cognition.2009.02.012

Ritter, E. (1993). Where's gender? *Linguistic inquiry*, *24*(4), 795–803. Retrieved from https://www.jstor.org/stable/4178843

Roca, I. M. (1989). The organisation of grammatical gender. *Transactions of the Philological Society*, *87*(1), 1–32.

Royen, G. (1929). *Die nominalen klassifikations-systeme in den sprachen der erde: historisch-kritische studie, mit besonderer berücksichtigung des indogermanischen* (No. v. 4). Mechitharisten-buchdruckerei. Retrieved from https://books.google.es/books?id=CYEtAAAAIAAJ

Ryan, K. (2010). Variable affix order: grammar and learning. *Language*, *86*(4), 758–791.

Saldana, C., Kirby, S., Truswell, R., & Smith, K. (2019). Compositional hierarchical structure evolves through cultural transmission: an experimental study. *Journal of Language Evolution*, *4*(2), 83–107. doi: 10.1093/jole/lzz002

Saldana, C., Oseki, Y., & Culbertson, J. (2021). Cross-linguistic patterns of morpheme order reflect cognitive biases: An experimental study of case and number morphology. *Journal of Memory and Language*, *118*, 104204. doi: 10.1016/j.jml.2020.104204

Santos, R. (1996). *Le mey: langue ouest-atlantique de guinée*. Unpublished doctoral dissertation, Université de la Sorbonne Nouvelle (Paris 3). (2 vols.)

Schadeberg, T. C. (2001). Number in swahili grammar. *Afrikanistische Arbeitspapiere 68: Swahili Forum VIII*, 7–16.

Seifart, F. (2005). *The structure and use of shape-based noun classes in miraña (north west amazon)*. Unpublished doctoral dissertation, Radboud Universiteit Nijmegen.

Skirgård, H., Haynie, H. J., Blasi, D. E., Hammarström, H., Collins, J., Latarche, J. J., ... others (2023). Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances*, *9*(16), eadg6175. Retrieved from https://grambank.clld.org/ doi: https://doi.org/10.1126/sciadv.adg6175

Smith, K. (2018). The cognitive prerequisites for language: insights from iterated learning. *Current opinion in behavioral sciences*, *21*, 154–160. doi: 10.1016/j.cobeha.2018.05.003

Stump, G. T. (1993). How peculiar is evaluative morphology? 1. *Journal of linguistics*, *29*(1), 1–36.

Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences*, *113*(16), 4530–4535.