

Linguistic Typology: Gender and Number

Gender and classifier systems

Tense is grammaticalized time; mood is grammaticalized modality. The extralinguistic category corresponding most closely to gender is sex, but in most languages which have gender, gender (masculine, feminine etc.) is much more than grammaticalized sex (male, female etc.). In fact, most languages which assign masculine gender to male entities and feminine gender to female ones have a certain amount of seemingly arbitrary gender assignment and sometimes even gender assignment which contradicts predictions one could make on the basis of sex.

Seemingly random gender assignment: in French, men are masculine (*il père* ‘the.MASC father’) and women are feminine (*la mère* ‘the.FEM mother’), but since gender assignment affects all nouns, most inanimate entities are assigned to one gender without regard to meaning. The following nouns are masculine: *drame* ‘tragedy, drama’, *gant* ‘glove’, *saphir* ‘sapphire’. The following nouns are feminine: *explosion* ‘explosion’, *grillade* ‘grilled meat’, *prière* ‘prayer’.

Gender assignment which is not in harmony with sex: in German, men are masculine (*der Mann* ‘the.MASC man’) and women are feminine (*die Frau* ‘the.FEM woman’), and there is a third gender called neuter (*das Haus* ‘the.NEUT house’). Some human beings are neuter: *Männlein* ‘little man’, *Fräulein* ‘Miss’, *Mädchen* ‘girl’, *Weib* ‘woman’ (archaic), *Kind* ‘child’. For some of these there is a simple rationale: nouns in *-lein* and *-chen* are diminutives, and diminutives are always neuter in German, regardless of sex; morphology overrides sex in German. *Fräulein* is only formally a diminutive, but not semantically, and for *Mädchen* the non-diminutive base noun *Maid* ‘young woman’ is distinctly archaic. Still, the nouns pattern with the frequent diminutives in *-lein* and *-chen*.

Indo-European languages with gender systems normally have two genders (masculine and feminine) or three genders (masculine, feminine, neuter). Other languages make similar distinctions between noun classes: Ojibwa (and other Algonquian languages) distinguish between ‘animate’ and ‘inanimate’ nouns. As with gender in Indo-European languages, such a distinction is partly semantic. The following are semantically animate and belong to the animate noun class: *enini* ‘man’, *enim* ‘dog’, *menito:* ‘Manitou’. The following are semantically inanimate and belong to the inanimate noun class: *essin* ‘stone’, *peka:n* ‘nut’, *wa:wan* ‘egg’. Some things Europeans tend to think of as inanimate are in the animate class, and this can be motivated by religious beliefs; trees for instance are animate. And there are certain nouns which are in the animate class and

for which it is difficult to find any such rationale, e.g. *meskomin* ‘raspberry’ and *uppwa:kan* ‘pipe for smoking’.

Other languages have more than three classes. Dyirbal has four, and Bantu languages normally have more than ten noun classes. If there are more than three such classes, people often speak of noun classes rather than genders, but the principles behind gender and noun classes are essentially the same.

An etymology and a formal definition of gender

Intuitively it makes sense to look at the systems of Indo-European languages in the same way as at the Algonquian systems, even though the semantic basis of the former is sex and of the latter, animacy. I shall refer to all such systems as gender systems. In non-technical usages, English ‘gender’ normally refers to sex-based classifications, but in typological studies an older meaning of the word is used.

The word ‘gender’ comes from Latin *genus*, a neuter noun meaning ‘kind’ or ‘sort’. I am using the English word in this sense. In Late Latin the noun became masculine and acquired a nominative / accusative *generem*, from where we get Old French *gendre*, and then English *gender*. French *genre* continues the Old French noun, and the second English loan, *genre*, is actually closer in meaning to Latin *genus* and the linguistic term *gender*. (Incidentally, German has also borrowed the French noun *genre* — as a neuter).

I shall treat different genders as a linguistic way of classifying nouns, as a way of sorting them. The definition of gender I adopt is a purely formal one and has nothing to do with sex; it goes back to Hockett (1958): ‘Genders are classes of nouns reflected in the behavior of associated words’. In other words, nominal genders are marked on adjectives, articles, verbs etc.

How random is gender?

To second-language learners genders often appear random. But it is unlikely that children acquiring a first language have to learn the gender of each and every noun separately. Native speakers agree on the gender of most nouns; exceptions are rare, e.g. German *Radio* ‘radio’ (neuter, but masculine in some dialects) or *Butter* ‘butter’ (feminine, but masculine in some dialects), and even here there is variation between only two genders rather than all three. Loan words are assigned genders automatically, again with practically all native speakers agreeing. *E-mail* for instance is a loan in German; since in English the pronoun used is *it* rather than *he* or *she*, prescriptive grammars will tell you that the noun is neuter, but most speakers actually use the feminine gender. So there must be criteria speakers use for assigning gender.

Semantic gender assignment

In some languages gender assignment follows semantic principles. In Tamil (Dravidian, South East India and Sri Lanka) there are three genders assigned by semantic principles:

- masculine gender: used for male rational entities; examples: *aan* ‘man’, *civaN* ‘Shiva’ (a Hindu god)
- feminine gender: used for female rational entities; examples: *pen* ‘woman’, *kaali* ‘Kali’ (a Hindu goddess)
- neuter gender: used for non-rational entities (including animals): *maram* ‘tree’, *viitu* ‘house’

Exceptions are rare and can be motivated: *cantiraN* ‘moon’ is masculine because it is also a deity. *Makavu* ‘child’ is neuter, but can be masculine or feminine if you want to be more specific. *Yaanai* ‘elephant’ is neuter, but can be masculine or feminine if you want to refer to a human with elephant-like qualities.

Dyirbal, a Pama-Nyungan language of north-east Queensland, has four genders marked on classifiers; gender I takes the classifier *bayi*, gender II takes *balan*, gender III takes *balam*, and gender IV takes *bala*. The system follows semantic principles, but allows for more exceptions than Tamil. Gender I is for male humans and non-human animates, e.g. most snakes and fishes. Gender II is for female humans, water, fire, and fighting. Gender III is for non-flesh food, and gender IV is for the remaining nouns.

Exceptions to this gender assignment are not random. They fall into three categories:

- a) mythological associations: birds, being non-human animates, ought to be in class I, but in mythology they stand for the spirits of dead human females and so are by and large in class II.
- b) concept association: fishing instruments, e.g. ‘fishing line’ and ‘fishing spear’, ought to belong to gender IV, but are associated with fish (gender I) and thus also belong to gender I.
- c) marking of an important property: gender II contains a number of dangerous things (fire, fighting). Fishes belong to gender I, but there are two dangerous species belonging to gender II, the stone fish and the gar fish.

Morphological gender assignment

In this section I shall present data from Latin and from Bantu languages. Both have a fairly complex noun morphology. Latin has five declension classes for nouns, which have distinct case and number allomorphs. The following table presents the nominative, accusative, and genitive singular endings of these five classes:

	Nominative	Accusative	Genitive
1st decl.	<i>-a</i>	<i>-am</i>	<i>-ae</i>
2nd decl.	<i>-us/-um</i>	<i>-um</i>	<i>-ī</i>
3rd decl.	various	<i>-em/-im</i>	<i>-is</i>
4th decl.	<i>-us</i>	<i>-um</i>	<i>-ūs</i>
5th decl.	<i>-ēs</i>	<i>-em</i>	<i>-eī</i>

Gender is marked by agreement on adjectives, participles and pronouns. Gender assignment is predominantly morphological: nouns of the first conjugation tend to be feminine, nouns of the second conjugation are masculine if the nominative has *-us* and neuter if the nominative has *-um*, nouns of the fourth declension are masculine, and those of the fifth are feminine. The third declension is a mixed bag and contains all three genders, largely depending on suffixes; e.g. nouns in *-tiō* are feminine and nouns in *-men* are neuter. Some third declension nouns do not have suffixes, and if these are rare words there is some synchronic gender variation.

There are some exceptions to morphological gender assignment:

a) nouns denoting humans are masculine or feminine, regardless of declension class. Very often such nouns are in classes which have the genders corresponding to sex anyway, but there are exceptions, and here sex-based gender assignment is more important than morphology (a pattern we find in language after language): *agricola* ‘farmer’ ought to be feminine because it belongs to the first declension, but is in fact masculine; *Astaphium* and *Pinacium* (Greek personal names) ought to be neuter because they are in the second declension and end in *-um*, but *Astaphium* is feminine and *Pinacium* is masculine. The only exceptions to sex-based gender assignment are *scortum* ‘prostitute’ and *mancipium* ‘slave’, both of which are neuter, as their declension class correctly predicts, because these humans are regarded as a commodity rather than as humans (but *meretrix* ‘prostitute’ is feminine because it is in the third declension and has a feminine suffix, and *servus* ‘slave’ is masculine because it is in the second declension and ends in *-us*). Note *socrus* ‘mother-in-law’, which is feminine despite belonging to the fourth conjugation; Italian *suocera* shows that the noun was eventually put into the first conjugation, which is predominantly feminine.

b) trees and cities are feminine, winds and rivers are masculine. Here we are dealing with tendencies that are not nearly as strong as the sex-based gender assignment under (a). Trees in second-declension *-us*, for example, are predominantly feminine (*pōpulus* ‘poplar’), but occasionally masculine, and anything ending in *-um* in the nominative is neuter, regardless of semantic associations. (Note that humans in *-um* are normally masculine or feminine!)

c) exceptions without rationale: these are rare, e.g. *porticus* (fourth declension, ‘portico’) is feminine, and so is *manus* ‘hand’ (also fourth declension). Sometimes such exceptions are regularized, e.g. *dies* is originally masculine (see related languages), but since all other fifth-declension nouns are feminine, there is variation in gender here, and the Romance reflexes are all feminine.

Bantu languages mark gender not just on associated words (adjectives, numerals, verbs), but also on the nouns themselves; compare standard examples from Swahili:

(1) *Kikapu kikubwa kimoja kilianguka*
ki-basket ki-large ki-one ki-fell
'One large basket fell.'

(2) *Vikapu vikubwa vitatu vilianguka*
vi-basket vi-large vi-three vi-fell
'Three large baskets fell.'

Bantu languages commonly have between ten and twenty gender prefixes, but *ki-* and *vi-* above would count as separate ones rather than as singular and plural; in effect, then, Bantu languages have between five and ten genders. For comparative purposes, Bantuists normally simply number these prefixes, which makes life difficult if you are looking at one single Bantu language instead of a group of languages; Swahili, for instance, has no classes 12, 13, or 14, but it does have class 15. *Ki-* is class 7 and *vi-*, its plural, is class 8.

A few things have to be noted. The first is that for some nouns the gender is not determined by the noun, but the lexical meaning is determined by the gender; e.g. *ki-ti*, plural *vi-ti*, is a 'wooden stool', but *m-ti*, plural *mi-ti*, is a 'tree'. This is because the prefixes have lexical meaning to some extent, even though most of it has been bleached out.

Now compare another Swahili example:

(3) *Kifarumdogo alikuwa hapa.*
ki-rhinoceros m-small m-was here
'A small rhinoceros was here.'

The prefix on the noun does not fit with the prefixes on the adjective and the verb. This phenomenon corresponds to the Latin example above: *agricola* 'farmer' belongs to the predominantly feminine first declension, but takes masculine agreement.

But let us finally look at gender assignment rules. Chichewa, another Bantu language, has a straightforward system:

a) animates, regardless of the noun prefixes, have semantic gender assignment: augmentatives ('giant snake' etc) belong to gender 5/6, diminutives belong to gender 7/8, and the remaining animates belong to gender 1/2. Semantics overrides morphological gender assignment (we saw this in the last Swahili example as well, where the rhinoceros was assigned gender by semantics rather than morphology).

b) other nouns have morphological gender assignment: nouns with 3/4, 5/6, 7/8, 9/10, 11/10, and 15 prefixes take adjectives, verbs etc. with 3/4, 5/6, 7/8, 9/10, 11/10, and 15 prefixes. Note that gender 15 does not have a plural because it is the gender assigned to infinitives only.

Phonological gender assignment

In some languages gender assignment is primarily determined by phonology, but as with morphological gender assignment there are normally nouns for which semantics overrides other criteria of gender assignment.

Yimas is a Papuan language with around 250 speakers. It was described in detail by W.A. Foley. Yimas has eleven genders which are marked by agreement on verbs and adjectives. Membership of genders I–IV is determined by semantic criteria: gender I is used for male humans ('man', 'father-in-law' etc.); gender II is used for female humans ('woman', 'mother' etc.); gender III is used for higher animals ('dog', 'crocodile'); gender IV is used for important plants ('sago palm', 'coconut palm').

Genders VI–XI are determined by phonological criteria:

Singular ending	Gender	Example	Gloss
- <i>ŋk</i>	VI	<i>kraŋŋk</i>	frog
- <i>mp</i>	VII	<i>impramp</i>	basket
- <i>i</i>	VIII	<i>awi</i>	axe
- <i>aw</i>	IX	<i>trukaw</i>	knee
- <i>uk</i>	X	<i>antuk</i>	mouth
- <i>uŋk</i>	XI	<i>awuŋk</i>	wall

Note that gender VI contains nouns in -*ŋk* with a vowel other than *u* before this final cluster; otherwise the noun is assigned gender XI.

Gender V is then used for all remaining nouns.

Gender assignment in French is generally considered to be arbitrary. Its parent language Latin had gender assignment which was by and large based on morphology, but most of the inflectional morphology of Latin was lost by sound changes, and many derivational suffixes were also lost or became opaque. The Latin three-gender system has in addition been reduced to a two-gender system.

However, more recent studies (Tucker, Lambert, and Rigault 1977) show that gender assignment in French is more or less predictable from the last phonemes. There are some sex-based rules which can override phonology: male humans are masculine and female ones are feminine, regardless of the ending; thus *père* 'father' is masculine and *mère* 'mother' is feminine, but the endings are the same. Otherwise, phonological rules apply.

Words can end in a huge variety of phoneme combinations, so I shall just give a few examples. Nouns ending in /ʒ/, e.g. *ménage* 'housekeeping', are masculine in 94.2% of all cases. Nouns in /z/ like *église* 'church' are feminine in 90% of the cases. Nouns ending in /ɛ/ are masculine in 90.2% of the cases, and if this vowel is nasal instead of oral they are masculine in 99% of the cases. Nouns ending in /p/ are masculine in 48.6% of the cases, so here the last phoneme is not enough to predict gender; if, however, you take into account the last two or three phonemes, gender becomes predictable again.

How well do these rules work? Tucker, Lambert, and Rigault state that phonological rules work in about 84.5%, which is already quite a good figure; if you consider that they counted as exceptional all cases where sex-based gender assignment overrides phonological rules, gender becomes highly predictable indeed.

Hybrid nouns and the agreement hierarchy

Hybrid nouns have properties of two genders. English has a few hybrid nouns, but since English by and large has a sex-based gender system marked only on pronouns, this may not be immediately obvious. But note that some nouns can be treated as neuter or feminine, for example, *ship* or *moon*, while others can be neuter or masculine, e.g. *sun*. Note incidentally that this hybrid treatment of *sun* and *moon* came about after the Old English gender system had collapsed; in this particular case the hybrid treatment is a poetic invention, while in Anglo-Saxon, just as in Modern German, *sun* was feminine and *moon* was masculine. Pets, for instance dogs and cats, are also often treated in hybrid fashion.

There are occasions when hybrid nouns have to be treated as belonging to one particular gender, i.e. occasions when there is no choice at all. The agreement hierarchy can help to determine when this is the case:

attribute < predicate < relative pronoun < personal pronoun

The more to the right an entity is on this hierarchy, the more likely it is to exhibit semantic rather than formal agreement. Let us look at English again: here gender is only marked on personal pronouns (*he, she, it*) and to some extent on relative pronouns (*who* masculine / feminine, *which* neuter). If you want to refer to your dog as a family member, masculine or feminine agreement is the semantic agreement, while neuter agreement is the formal agreement. If you treat your dog as a family member, you will use *he* or *she*, but you may well use *which* instead of *who*. If you personify a ship, you will say *she*, but the relative pronoun is still *which*.

In Latin many animals have masculine or feminine gender regardless of sex — it is just the declension class that matters for gender assignment. But what if sex matters, e.g. if you are discussing female animals as opposed to male ones? There is no clear strategy:

(4) ... *solēre elephantum gravidam perpetuōs decem esse annōs*. (Plaut. *Stich.* 168-9)

‘... that an elephant (second declension) is normally pregnant (feminine) for a whole ten years.’

(5) *Quī lepus dīcitur, quom praegnās sit, tamen concipere*. (Varro *rust.* 3. 12 5)

‘This kind (masc.) of hare (third declension) is said to conceive even when it is pregnant.’

Elephantus in Ex. 4 would normally take masculine adjectives, but is combined with a feminine form to show that it refers to a female animal. In Ex. 5, on the other hand, *lepus* goes with a masculine pronoun, even though it also refers to a female animal. Note that the adjective in Ex. 4 is part of the predicate, while the pronoun in Ex. 5 is attributive, so that the agreement hierarchy is not violated.

Number

Number can be marked on nouns, which is rare for gender (but cf. the Bantu data above). It can also be marked on pronouns, adjectives, verbs, and participles. Most languages have some kind of number marking, at least on pronouns. Cantonese for example does not distinguish number among nouns, but clearly among personal pronouns. Thus, *hēungjǔu* can refer to one, two, or more bananas (and classifiers offer no help here), but plural personal pronouns have a suffix that distinguishes them from singular ones:

	Singular	Plural
1st person	ngóh	ngóh-deih
2nd person	néih	néih-deih
3rd person	kéuih	kéuih-deih

This is not quite as self-explanatory as it looks because the first person plural, despite its name, is of course not just a plural of the singular in semantic terms; *we* can mean ‘I and you’, in which case we are not dealing with a plural of speakers, but with one speaker and one addressee.

Pirahã, the only remaining member of the Mura family, is spoken in the Amazonas region; the language has no number marking, not even on pronouns.

Just as tenseless languages are not deficient because time can of course be expressed by lexical means, languages without number can express numerals by lexical means. In fact, number in English is a fairly uninformative category; you only distinguish between one and more than one, but not for example between two and three.

The number hierarchy

Singular and plural are not the only numbers attested. Some languages have a dual for two items, a trial for three, or a paucal for a few. In earlier works you could find the following number hierarchy:

singular > plural > dual > trial / paucal

This means that if a language has two numbers, it will be singular and plural. If it has three numbers, it will be singular and plural and dual. Note that a language with only one number can strictly speaking not be argued to have singulars only, as this one number will be used for everything.

We are dealing with a tendency here, not an absolute hierarchy. Bayso, a Cushitic language spoken in Ethiopia, does not fit in: it has a singular, a plural, and a paucal. The number hierarchy has to be modified; if there are two or three numbers, the hierarchy is as follows:

singular > plural > other

If there are three or more numbers, it is as follows:

singular > plural > dual > other.

We can find the following patterns: English has singular and plural; some forms of Ancient Greek have singular, plural, and dual; Bayso has singular, plural, and paucal; Larike, a Moluccan language (part of the Austronesian family), has singular, plural, dual, and trial; and Yimas, a Papuan language, has singular, plural, dual, and paucal.

The hierarchy is not just an inventory of what numbers a language can have, but also important for other reasons. Sometimes number marking is not obligatory. In Vedic Sanskrit you have to use duals for two members, but in Slovene, which has a dual, this is not necessary:

- (6) Nóge me bolijo.
foot.PL 1SG.ACC hurt.PL
'My feet hurt.'

Now the hierarchy tells us that if number marking is not obligatory, it will be facultative among the numbers at the right end of the hierarchy rather than at the left end.

The animacy hierarchy

I introduced the animacy hierarchy in connection with split ergativity:

1st / 2nd person pronouns > 3rd person pronouns > kinship terms > terms for humans > animate common noun > inanimate common noun

The items to the left are more likely to follow the nominative-accusative pattern than the ones to the right because the ones to the left tend to be more topical.

The items to the left are also more likely to have more number distinctions than the items to the right; or the items to the left have obligatory number distinctions where the items to the right have facultative number distinctions; or if number is facultative everywhere, it tends to be marked more frequently among the items on the left.

Compare Slave, an Athabaskan language; Slave marks plural by a suffix *-ke*, but this marking is restricted to humans and dogs, e.g. *t'eere* 'girl', *t'eere-ke* 'girls'.