Chapter 3

Gender: esoteric or exoteric?

Östen Dahl

Stockholm University

Although grammatical gender would seem to be a paragon example of a mature phenomenon in the sense of Dahl (2004), it turns out to be hard to establish any correlation to ecological parameters that have been claimed to co-vary with other such phenomena, such as community size and degree of contact. Grammatical gender also does not seem to correlate with morphological complexity in general. Our understanding of these relationships is hindered by the areal and genetic skewings in the distribution of gender and the lack of diachronic data. To understand how the ecological factors influence the growth, maintenance, and demise of gender systems and eventually their synchronic distribution, we have to go beyond the patterns that can be found in typological data bases like *WALS*. In particular, we need to know more about the conditions under which gender systems arise and mature.

Keywords: grammatical gender, esoteric niche, exoteric niche, language ecology, morphological complexity, mature phenomenon, areal typology, community size, suboptimal transmission, semantic gender assignment, formal gender assignment.

1 Introduction: The esoteric-exoteric distinction and morphological complexity

In recent decades, many authors have suggested that there is a connection between grammatical complexity, in particular morphological complexity, and factors external to the language system, such as community size, the degree of contact with other language communities and the extent to which the language is learnt and used by non-native speakers (see e.g. the discussion in Trudgill 1983 and Dahl 2004). It appears obvious that a language with grammatical gender is ceteris paribus more complex than one without grammatical gender, but can we

say anything about the relationship between grammatical gender and the "ecology" of the language, that is, the conditions under which it is used, learnt and transmitted to new users?

Over the last ten years, there have also been attempts to study the relationship between grammatical complexity and language ecology by quantitative methods. Thus, Sinnemäki (2009: 138) finds in a cross-linguistic investigation that there is "a statistically relatively strong association between community size and complexity in core argument marking, measured as adherence to versus deviation from the principle of one-meaning—one form". In another study, Lupyan & Dale (2010) make a distinction between "languages spoken in the esoteric niche", i.e. languages with comparatively smaller populations, smaller areas, and fewer linguistic neighbours, and those spoken in the "exoteric niche", i.e. languages with larger populations, larger areas, and more linguistic neighbours. Basing themselves on data from the *World atlas of language structures* (*WALS*; Dryer & Haspelmath 2013), they list more than a dozen morphological features which they have found are more common in languages spoken in the esoteric niche:

- case markings

- ergative alignment

- grammatical categories marked on the verb

- person marking on adpositions

- noun/verb agreement

- inflectional evidentiality

- affixal negation

- morphological future tense

- remoteness distinctions in the past tense

- alienability/inalienability distinctions

- optative mood marking

- distance distinctions in demonstratives

- morphological marking of pronominal subjects

- separate associative plurals

An earlier work that also should be mentioned here is Perkins (1992), who found a negative correlation between language complexity as manifested in deictic grammatical distinctions and cultural complexity as measured by a variety of factors, including the size of communities.

2 Is grammatical gender correlated with esotericness?

In Dahl (2004), I introduced the notion of MATURITY as applied to grammatical phenomena. A grammatical pattern was said to be mature if it has a non-trivial prehistory in any language where it appears. I argued that in situations of "suboptimal transmission" of languages, mature patterns will be transmitted less easily and will tend to be reduced or eliminated. As one of "the most mature phenomena in language", I pointed to grammatical gender. The kind of gender systems we see in some of the major European languages arguably passed through a number of intermediate stages before becoming what they are today. Gender is also a category that depends on inflectional morphology and is conspicuously absent from languages that lack it, such as creoles and the isolating languages of South East Asia and West Africa. We would therefore expect gender to be among the features that have a negative correlation with language size and a positive correlation with general morphological complexity.

But it turns out to be surprisingly difficult to find any such correlation. Already Perkins (1992: 157) points to gender in pronouns and verb affixes as lacking the clear negative correlation with cultural complexity that he finds with other grammatical features such as deictic distinctions in demonstratives. Similarly, gender is not among the features listed by Lupyan & Dale (2010) as being correlated with their esoteric/exoteric dimension. Gary Lupyan (personal communication) informs me that while no consistent relationship can be found between population and sex-based gender systems in the data from *WALS*, there is a weak positive correlation between non-sex-based gender and population, that is, the opposite to what could be expected from what has been said above.

I have made some calculations of my own on the data in the three *WALS* chapters on gender systems (Corbett 2013a,b,c), using iterated samples of one language from each of 60 families or 100 genera, and computing the mean and median values for Pearson's *r* correlating those samples to the logarithm of the number of speakers of each language (using figures from the Ethnologue). This essentially confirmed the findings of Lupyan and Dale, including the weak positive correlation for non-sex-based gender.[1]

---

[1]In Dahl (2011), I reported a positive correlation (0.142) between number of genders and number of speakers in the *WALS* data. That calculation was done on the whole sample, however, and thus did not take account of possible areal and genetic biases.

It is questionable if any firm conclusion can be drawn from the last finding. Judging from the data in *WALS*, non-sex-based gender systems are relatively uncommon – Corbett (2013c) classifies 28 out of 112 gender systems (in a sample of 257 languages) as belonging to this type, and of these 18 are from one single family (Niger-Congo). The total number of families where languages with non-sex-based gender are found is seven, which in my view makes the number of independent cases too small to draw any conclusions.

Thus, we can conclude that it is not possible to show from the data at hand that the presence of gender – or specific types of gender – is correlated to ecological factors such as population. Rather, the evidence suggests the absence of any correlation in any direction (or possibly a very weak positive one).

3 Grammatical gender and morphological complexity

I said above that everything else being equal, a language with grammatical gender is more complex than one without grammatical gender. It does not follow, however, that gender is correlated with other kinds of complexity. In fact, Nichols (2019 [this volume]) argues on the basis of a sample of 146 languages that there is no significant difference between gender languages and genderless languages in (i) overall complexity; (ii) morphological complexity in general; (iii) degree of inflectional synthesis of the verb.

These findings can be seen as being in line with the lack of a correlation between gender and ecological factors in the sense that a connection between those factors and a large number of features involving morphological complexity has been demonstrated. On the other hand, the findings are puzzling since gender – following Corbett (1991: 4) – is by definition realized as agreement, and agreement, or perhaps better indexation, would normally be manifested in inflectional morphology. Accordingly, gender is not found in languages traditionally classified as isolating, as noted above.

Trying to elaborate on Nichols' findings, I looked for a correlation between gender and any specific inflectional category in the *WALS* data, but did not find anything close to significance, not even with nominal categories such as case and number. Given that gender and number often go together in inflectional systems, the last finding is particularly puzzling. However, the situation is different if we look just at the languages that have both "semantic and formal gender assignment" and plural marking. For the 26 languages in this group for which there is also information on plural marking, 25 have a morphological plural and out of these, 23 languages mark plural obligatorily on all nouns. In other words, if a

language has gender with formal assignment, it will also tend to have a highly grammaticalized nominal number system.

4 Areal and genetic skewings in the distribution of gender

What is easily seen in the *WALS* material is that there are strong skewings in the geographical distribution of gender. About two thirds of the languages with gender systems in Corbett's sample are from Africa and Eurasia; the percentage of gender languages among the languages from those continents is 59, compared to 30 in the languages from the rest of the world. Particularly striking is the distribution of languages with "semantic and formal gender assignment" (Corbett 2013c), where as many as 53 of 59 are found in Africa, Europe, and south-western and southern Asia. Furthermore, nearly all these languages belong to three large families – Afro-Asiatic, Indo-European, and Niger-Congo, which also happen to contain many languages with high speaker numbers, and the few remaining languages are either Nakh-Daghestanian or Khoisan.

In view of what was just said, it would be desirable to factor out possible areal influence from the calculations. This however meets with the problem that the ecological factors that we would like to correlate with the presence of gender are geographically skewed to the same degree, and, in fact, in a similar way. Thus, while 53 of the languages from Africa and Eurasia in Corbett's sample have more than a million speakers, there is just one such language (Guaraní) representing the rest of the world (Australia, the Pacific and the Americas). A more generous sampling would turn up a few more, but it would hardly change the general picture. Nevertheless, it is of some interest to see what happens if the languages from Africa and Eurasia are removed from the calculations of correlation. The results differ only marginally from the ones obtained from the total sample, however, and again it may be questioned if the sample isn't simply too small.

The general conclusion seems to be that it is hard to correlate gender to anything at all, at least as long as we restrict ourselves to the data in *WALS*. It would clearly be better to have a larger sample, but it is not obvious that it would help in the end, due to the heavy areal skewings we find both in gender systems and in the ecology of languages.

5 The diachronic perspective

Another problem is the limitation to synchronic data. One observation is that the clustering of gender languages in western Eurasia and adjacent areas of Africa

actually grows stronger as we go back in time and the area occupied by the in-volved families shrinks. Levins (2002: 252) argues that the Indo-European distinction between masculine and feminine probably arose under Semitic influence, and Matasović (2012) thinks that Indo-European may have influenced those Caucasian languages that have genders. In any case, we cannot unreservedly treat the gender systems in Indo-European, Semitic and Nakh-Daghestanian as independent developments.

In this context, it is important to remember that the probability that a given language exhibits a grammaticalized pattern will depend at least on two different parameters: the propensity for the pattern to arise and the propensity for it to be eliminated in one way or another. It has been claimed (e.g in Dahl 2004: 199) that gender systems are very stable. What we can see in Corbett's sample is that the families in the western Old World where gender systems with formal assignment show up are very homogeneous as to the presence of gender. Looking at the languages of western Europe, one gets the impression that gender is among the last categories to go when a language undergoes general morphological simplification; thus, many Romance and Germanic languages have lost their case systems but kept gender, although in a somewhat reduced form. It is somewhat hard to generalize here, however – Armenian is an example of a language which has lost gender but preserved its case system (see e.g. Kulikov 2006). It can also be difficult to decide if a category has really disappeared – there may be remnants such as the s-genitive in the Germanic languages, or there may be a renewal of a system, as in the Indic languages, where new case systems have appeared. There is no doubt, however, that a gender system may take a long time to develop but that once it has arisen, it can continue to exist for a very long time. This is bound to weaken the synchronic connection between the presence of gender and ecological factors such as population size, as a gender system may be preserved even if the external situation of the language changes. Moreover, although it is well known that gender systems tend to break down in situations of suboptimal transmission, as in creolization, we know less about the ecological conditions that favour the rise of gender systems.

6 Developing the typology of gender systems

It is thus likely that we have to go beyond synchronic typology to arrive at a fuller understanding of the relationship between gender systems and ecological factors. Detailed comparisons of developments within one and the same family (along the lines of Di Garbo & Miestamo 2019 [in Volume II]) may shed light

on the problem. But we may also need a more elaborate typology of gender systems, for instance by taking into account in a more systematic way the domains where they operate, and also sharpen the definitions of the features currently used to classify gender systems. Thus, we saw above that the gender systems that are labelled as having "semantic and formal gender assignment" both had a specific geographical distribution and a high correlation with highly grammaticalized grammatical number. On the other hand, the classification behind this label is not fully understood. Corbett (1991: 62) notes that in languages with formal assignment of gender, the gender of a noun is often "evident from its form", and calls this "overt gender", as opposed to "covert gender". He says that in an ideal overt system would have "a marker for gender on every noun" and mentions Swahili as an example of a system that approaches this ideal. But this raises the question of what is basic – the marker or the gender. In fact, the borderline between marking gender and being the source of it is quite thin. For Bantu languages to have overt gender it is necessary to consider the prefixes as being parts of nouns. But consider now Khasi (Austroasiatic), which is treated as having semantic gender assignment in Corbett (2013c). In Khasi, nouns are obligatorily preceded by a "pronominal marker". There are four such markers: *u* masculine, *ka* feminine, *i* diminutive and *ki* plural. The same elements show up as obligatory 3rd person subject markers. Nagaraja (1985: 7) says that "[a] noun without a pronominal marker is not possible" but still treats combinations of pronominal markers and nouns as two-word phrases, in order to "facilitate the dealing with the structure of the nouns as such". If this choice had not been made, Khasi would look as having a mini-version of a Bantu noun class system, with "overt gender". We meet a rather similar problem in trying to draw a distinction between gender marking and inflectional classes, as argued in Dahl (2000), exemplified by Scandinavian definite articles, which are manifested both as independent words and as suffixes on nouns, but which vary according to gender in a uniform way wherever they occur (see Dahl 2000 for a discussion).

If we question the role of morphemes such as Bantu noun prefixes as the source of gender assignment, we may also have to reconsider the view that gender assignment is generally rule-governed. Both Killian (2019 [this volume]) and Svärd (2019 [this volume]) argue for the significance of "opaque" or "arbitrary" gender, a possibility that has been downplayed in recent decades. It may be noted that the rise of opaque gender assignment can be seen as an indication of the maturity of a gender system, since it is likely to appear at a relatively late stage of development.

7 Conclusion

Although grammatical gender would seem to be a paragon example of a mature phenomenon in the sense of Dahl (2004), we have seen that it is very hard to establish any correlation to parameters that have been claimed to co-vary with other such phenomena. To understand how the ecological factors influence the growth, maintenance, and demise of gender systems and eventually their synchronic distribution, we have to go beyond the patterns that can be found in typological data bases like *WALS*. In particular, we would need to know more about the conditions under which gender systems arise and mature.

References

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

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

Corbett, Greville G. 2013b. Sex-based and non-sex-based gender systems. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/31.

Corbett, Greville G. 2013c. Systems of gender assignment. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/32.

Dahl, Östen. 2000. Elementary gender distinctions. In Matti Rissanen, Terttu Nevalainen & Mirja Saari (eds.), *Gender in grammar and cognition*. Vol. 2: *Elementary gender distinctions: Manifestations of gender* (Trends in Linguistics. Studies and Monographs 124), 577–593. Berlin: de Gruyter.

Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.

Dahl, Östen. 2011. Are small languages more or less complex than big ones? *Linguistic Typology* 15(2). 171–175.

Di Garbo, Francesca & Matti Miestamo. 2019. The evolving complexity of gender agreement systems. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume II: World-wide comparative studies*, 15–60. Berlin: Language Science Press. DOI:10.5281/zenodo.3462778

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/.

Killian, Don. 2019. Gender in Uduk. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 147–168. Berlin: Language Science Press. DOI:10.5281/zenodo.3462764

Kulikov, Leonid. 2006. Case systems in a diachronic perspective: A typological sketch. In Leonid Kulikov, Andrej L. Malchukov & Peter de Swart (eds.), *Case, valency and transitivity*, 23–47. Amsterdam: John Benjamins.

Levins, Saul. 2002. *Semitic and Indo-European.* Vol. 2: *Comparative morphology, syntax and phonetics.* Amsterdam: John Benjamins.

Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PLOS one* 5(1). e8559.

Matasović, Ranko. 2012. Areal typology of Proto-Indo-European: The case for Caucasian connections. *Transactions of the Philological Society* 110(2). 283–310.

Nagaraja, Keralapura Shreenivasaiah. 1985. *Khasi, a descriptive analysis.* Poona: Deccan College Post-Graduate & Research Institute.

Nichols, Johanna. 2019. Why is gender so complex? Some typological considerations. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 63–92. Berlin: Language Science Press. DOI:10.5281/zenodo.3462760

Perkins, Revere D. 1992. *Deixis, grammar, and culture.* Amsterdam: John Benjamins.

Sinnemäki, Kaius. 2009. Complexity in core argument marking and population size. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 126–140. Oxford: Oxford University Press.

Svärd, Erik. 2019. Gender in New Guinea. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 225–276. Berlin: Language Science Press. DOI:10.5281/zenodo.3462770

Trudgill, Peter. 1983. *On dialect: Social and geographical perspectives.* New York: New York University Press.