Remarks on the complexity of gender

Danniel da Silva Carvalho*

Abstract

This squib presents some aspects of the nature of gender complexity and how it can be conceived throughout the linguistic literature. Emphasis is placed on the empirical difficulties that need to be considered in assessing the nature of gender as a category in grammar. Our target is to discuss the morphosyntactic manifestation of gender as a (complex) feature, focusing on the difficulty of defining it as a traditional feature. As gender can be defined as a system of agreement classes (CORBETT, 1991), agreement is the criterion used to decide how many genders a language has and which nouns belong to which gender. We present current proposals about formal and functional gender manifestation throughout languages, aiming at describing their behaviour and possible consequences for the syntax of this category. According to Kramer (2015), natural/biological and arbitrary gender have the same pre-syntactic nature, with the derivation responsible for differentiating them post-syntactically. However, the classification of such category as biological and/or arbitrary brings along with it important discussions on how this process determines the distribution and the semantic-pragmatic interpretation of such structure, taking into account, for instance, the speaker's world knowledge. This complexity, nonetheless, can be interpreted by means of the computation of gender.

Keywords: Gender, Grammatical category, Feature complexity

Resumo

Este *squib* apresenta alguns aspectos da natureza da complexidade de gênero e como isto pode ser concebido ao longo da literatura linguística. Damos ênfase em dificuldades empíricas que precisam ser consideradas ao se acessar a natureza de gênero como uma categoria na gramática. Nossa meta é discutir a manifestação morfossintática de gênero como um traço (complexo), com foco nas dificuldades em defini-lo como um traço tradicional. Visto que gênero pode ser definido como um sistema de classes de concordância (CORBETT, 1991), concordância é o critério usado para se decidir quantos gêneros uma língua tem, e quais nomes pertencem a cada gênero. Mostramos algumas propostas correntes sobre a manifestação formal e funcional de gênero nas línguas, com o objetivo de descrever seu comportamento e possíveis consequências para a sintaxe desta categoria. De acordo com Kramer (2015), gêneros natural/biológico e arbitrário têm a mesma natureza

---
*Universidade Federal da Bahia, UFBA. *E-mail*: dannielcarvalho@ufba.br.

pré-sintática, sendo a derivação a responsável por diferenciá-los pós-sintaticamente. Entretanto, a classificação de tal categoria como biológica e/ou arbitrária levanta uma importante discussão sobre como este processo determina a distribuição e a interpretação semântico-pragmática de tal estrutura, levando em conta, por exemplo, o conhecimento de mundo do falante. Esta complexidade, todavia, pode ser interpretada por meio da computação de gênero.

Palavras-chave: Gênero, Categoria gramatical, Complexidade de traços

1   Remarks on gender

Because this paper is an investigation on the behaviour of gender, it is necessary to properly define what we are calling *gender*. One traditional definition of gender is given by Kramer (2015, p. 109):

(1)   **Definition of Gender**

Gender is the sorting of nouns into two or more classes, as reflected in agreement morphology on determiners, adjectives, verbs and other syntactic categories.

According to Kramer, the definition in (1) has three main parts: there must be at least two genders, gender is restricted to nouns, and genders (nominal sub-categories) are differentiated from one another via agreement patterns.

The minimum of two genders is simple to justify: if all nouns agree in the same way, then there is no need to describe the noun as anything other than a single category defined by sheer noun-hood, not as sub-classes. As for the restriction to nouns, it is possible to identify sub-classes of verbs and adjectives (as well as other syntactic categories) based on varying linguistic behaviour in certain contexts (e.g., a sub-class of verbs takes a certain set of inflections, predicative adjectives behave differently than non-predicative adjectives in copular clauses). However, their differences do not manifest in different agreement patterns on other elements because nouns are (usually) the only category that controls (or triggers) agreement (COBERTT, 2006, p. 13).

Thus, even assuming that agreement patterns identify a gender system and differentiate gender in languages, it is not enough to morphologically specify a noun for its gender. The contrast in (2) and (3) shows that the agreement system is not enough to determine a particular gender in a noun according to its agreement patterns:

(2)   FRENCH

    a. la   petite   vache
       the.F small.F cow

     b. le   petit   taureau
       the.M small.M bull

(3)   HUNGARIAN

     a. a   kicsi tehén
       the small cow

     b. a   kicsi bika
       the small bull

(KRAMER, 2015, p. 110-111)

As the author notes, in Hungarian, 'cow' and 'bull' are distinct words and carry a component meaning of biological sex, but since there is no agreement pattern, the language is considered to have no gender system.

The terms *gender class* and *gender languages* need to be defined more precisely and with more explicit reference to the framework of nominal classification. While in numeral classifier systems the class membership of nouns is marked only in restricted syntactic contexts (mainly in matter of quantification), class membership in noun class languages triggers agreement on a range of elements inside and outside the noun phrase. Noun class languages have a relatively small number of classes (little more than 20, according to Hellinger and Bußmann, 2001). These classes consistently structure the entire nominal lexicon. Hellinger and Bußmann suggest, then, that there must be a differentiation between *gender languages* and *noun class languages*, based on grammatical and semantic facts. This distinction is also motivated by the paramount interest of this paper in the representation of *feminine/female* and *masculine/male*.

Gender languages are illustrated by many Indo-European and Semitic languages. These languages have only a limited number of *gender classes*, usually two or three (HELLINGER and BUßMANN 2001, p. 5). Nouns do not necessarily carry markers of class membership, but there is (obligatory) agreement with other word classes, both inside and outside the noun phrase. Most importantly, class membership is anything but arbitrary in the field of animate/personal reference, even with the so-called *arbitrary gender*, as shown by Kramer (2015). Languages of this type are traditionally called *gender languages* or *languages with grammatical gender*. A large number of languages belong to this group: Arabic, Brazilian Portuguese, Czech, Danish, Dutch, French, Greek, Hebrew, Hindi, Icelandic, Italian, Norwegian, Polish, Romanian, Russian, among many others.

Noun class languages display no obvious correspondence between class membership and a noun's specification as female-specific or male-specific. These languages, like Swahili, for instance, have a larger number of classes than gender languages. Often class membership is explicitly marked on the noun, like prefixes in Bantu languages, for example, which triggers extensive agreement on other word classes.

The lack of grammatical gender in a language does not mean that this category cannot

be communicated. There are various other means of representing gender (e.g. *lexical* or *social* gender), which can be employed to transmit gendered messages. Nevertheless, all these categories for gender — grammatical, lexical, referential and social, as presented in Hellinger and Bußmann (2001) — in these languages have a grammatical counterpart. Any representation of gender either as a morphological, a semantic or a pragmatic category demands a proper reading within the computation in any of the aforementioned languages.

2   The manifestation of gender features

Gender then can be defined as a morphosyntactic feature, since it is relevant to both morphology and syntax. Compared with person and number, it is the most puzzling feature that has raised the most questions in the linguistic literature (AUDRING, 2009, 2014; CORBETT, 1991, 2006, 2013; CURZAN, 2003; DUKE, 2010; KRAMER, 2009, 2015; MILLS, 1986; MOORE, 1921; SIEMUND, 2008; TRUDGILL, 1999, among many others). In the literature, the Saussurean linguistic sign is a pairing of form and meaning. In this view, each piece of form should be functional, each function should be expressed only once in an utterance and the relation between form and function should be systematic, either in terms of lexical meaning or grammatical function. A linguistic system with these properties is often called *transparent* (HENGEVELD, 2011). Recently, it has been assumed that transparency is associated with language in early stages of development (see AUDRING, 2014). As systems progress through their life cycle, they usually develop an increasingly degree of opacity.

According to Audring (2014), gender systems defy three criteria for transparency. Firstly, grammatical gender is highly irregular in its functionality. Gender information in a sentence contributes very little to the informational value of an utterance. The author shows the contrast between English and Italian:

(4)   a. a long story

      b. un-**a** lung-**a** storia   (AUDRING, 2014, p. 6)

Both sentences have the same meaning, saving the additional grammatical information in (4b), since Italian, as defined above, is a gendered language and marks gender redundantly throughout the sentence (as seen in boldface). Why languages grammaticalize such markings is not well understood.

Secondly, gender is formally complex. The two characteristics responsible for this complexity are inherent to all morphosyntactic features: redundancy and displacement. Redundancy stands for higher complexity because it defies the one-to-one mapping of form and function. Moreover, the information expressed is displaced in that "one word [carries] the grammatical meaning relevant to another" (CORBETT, 2006, p. 1). Still according to Audring

(2014, p. 6), the gender information belongs to the noun, yet it is expressed on other words elsewhere.[1] The fact that gender appears overwhelmingly elsewhere and, in some languages, it simply does not appear, even in the noun, illustrates its inherent complexity.

Finally, there is the question of whether gender features have or do not have semantic content. There are two major semantic domains often mentioned in relation to gender: sex and animacy. Yet, many gender systems are not straightforwardly or obviously semantic. Corbett (1991, p. 34) claims that all gender systems have a "semantic core". However, this core is often surrounded by a network of rules that make reference to other properties (AUDRING, 2014, p. 7). Audring (2014) assumes then that gender system is a complex, based on three dimensions: (i) the number of gender values; (ii) the number and the nature of assignment rules; and (iii) the amount of formal markings.

Regarding (i), the author claims that it is the most obvious criteria, since this constitutes a case of *constitutional complexity* (see RESCHER, 1998), the logical minimum number of genders being two. Two-gender systems are the most common variant. According to The World Atlas of Language Structure (WALS), Corbett (2013), from a sample of 257 languages, 112 maintained or have developed gender systems. From this languages, 50 (45%) are two-gendered. Three-gender systems are almost half (26 examples), and four-gender systems are about half of that (12). Yet, the languages with the largest gender systems identified so far are: Mountain Arapesh, with 13 genders; Ngan'gityemerri, with 15 genders; and Nigerian Fula, with around 20 genders (CORBETT, 2013 for references).

As gender is defined as a system of agreement classes, agreement is the criterion by which it is traditionally decided how many genders a language has and which nouns belong to which gender. Thus, considering (5), from Italian, we know that *donna* (women) is feminine and *uomo* (man) is masculine, because they take masculine and feminine agreement, respec-

---

[1]It is controversial that gender information is a prerogative of the noun. For instance, Cyrino, Armelin and Minussi (2015, p. 78) show some data from Brazilian Portuguese in which there is gender mismatch between the determiner and the noun, as in (i) below:

(i)   a. **A**   **garrafa**   está na minha casa.
         the.F bottle.(F) is   in my.(F) house.(F)

         'The bottle is in my house'

     b. **O**   **garrafa**   está na minha casa.
         the.M bottle.(F) is   in my.(F) house.(F)

         'Bottle is in my house'.

(ia) would accept another interpretation, similar to (ib) (i.e., "Bottle is in my house"), in which *Bottle*, a proper noun, would be interpreted as a female individual (i.e., "A girl whose nickname is 'bottle' is in my house"). On the other hand, the opposite picture is not allowed, since *bottle*, in (ib), can only refer to a human referent due to the use of the masculine definite article ('o' – 'the'). Therefore, according to the authors, gender information should be accounted for as sets within the Encyclopedia and, then, gender licensing is taken as an epiphenomenon of Encyclopedic interpretation.

tively. The final vowel of the noun, by contrast, is a reliable indicator of gender. *Mano* (hand), in (5d), on the other hand, looks like a masculine noun, but it is in fact feminine, as agreement shows. However, we can examine any agreement target and arrive at the conclusion that the indefinite article *una* and the adjective *alta*, in (5a), are equally good evidence for the feminine gender of *donna*, as the definite article *la* and the adjective *vecchia*, in (5c).

(5) a. un-**a**   donna   alt-**a**
         INDEF-F.SG WOMAN(F).SG TALL-F.SG
         'a tall woman'

     b. **un**   uomo   alt-**o**
         INDEF-M.SG MAN(M).SG TALL-M.SG
         'a tall man'

     c. l-**a**   donna   vecchi-**a**
         DEF-F.SG WOMAN(F).SG OLD-F.SG
         'an old woman'

     d. l**a**   mano   sinistr-**a**
         DEF-F.SG HAND(F).SG LEFT-F.SG
         'the left hand'

However, not all gender systems are straightforward as the Italian data above suggests. Four conditions must be met for the agreement method to be fail-safe:

 (i) controllers and targets distinguish the same gender value;

 (ii) all targets distinguish the same gender value;

 (iii) controllers are constituents in the gender value they trigger on a particular target;

 (iv) given the same controller, all targets show the same gender in all circumstances.

As Corbett shows (1991, p. 150), not all languages adhere to these principles. Romanian famously violates the first condition by having either two or three genders, depending on the perspective. On the perspective of the controller, it has three: nouns need to be sorted into three classes in order to explain their effect on agreement elements. Yet, the agreeing elements themselves only have morphological markers for two different genders.

	Singular	Plural
Masculine	*-∅*	*-i*
Feminine	*-ă*	*-e*

Table 1. Romanian gender morphology (AUDRING, 2014, p. 8)

According to Audring (2014), the fact that Romanian has to take agreement from the masculine paradigm in the singular and from the feminine paradigm in the plural necessitates the postulation of a third gender. The mismatch between controller and target genders has triggered an extensive debate around the question of whether Romanian has two or three genders. However, Audring assumes that something like Table 2 is more suitable to describe the gender paradigm in Romanian. The only particularity is that the patterns of syncretism are such that there are no agreement markers unique to neuter gender.

	Singular	Plural
Masculine	∅	*-i*
Feminine	*-ă*	*-e*
Neuter	∅	*-e*

Table 2. Romanian genders reconsidered (AUDRING, 2014, p. 8)

Audring points out that a similar problem arises when the target fails to mark the same array of gender, which is very common in Germanic languages.

The second dimension in which gender can be more complex concerns the system of assignment rules. In an ideal language, assignment of agreement rules are expected to be straightforward and semantic. This is not an unreasonable assumption, given that gender systems, as far as we know, arise from classifier systems or else from pronominal reference, which are both semantic in nature. Additionally, less than half of the 112 gender languages in the WALS' sample are exclusively organized on a semantic basis. The majority employs a mixture of formal and semantic rules, whereby formal rules may refer to phonological and/or morphological properties of the noun. Each sort of rule is exemplified below in (6):

(6)   a.   nouns referring to female humans are feminine;

b.   nouns ending in an accent vowel are feminine;

c.   nouns derived by means of the suffix *-tion* are feminine.

Thus, (6a) is a semantic gender assignment rule common in the world's languages; (6b) is a phonological rule described for Qafar, an East Cushitic language (see CORBETT, 1991, p. 51); and (6c) is a current morphological rule in German.

This type of rule does not say much about the complexity of the rule system, but these kinds of rule differ substantially in generality and thus in the number of nouns they can account for. Audring (2014) proposes a distinction between *large rules* (i.e., rules that are large in scope), and *small rules* (i.e., rules that are small in scope).

(7)   a. nouns denoting male persons are masculine, all others are neuter;

        b. nouns ending in an accented vowel are feminine, all others are masculine;

        c. nouns belonging to declension class I are masculine.

(8)   a. nouns denoting domestic trees are feminine in Scandinavian;

        b. nouns denoting "functional hollows" are neuter in German;

        c. nouns denoting cars are feminine in Italian;

        d. nouns denoting phenomena that naturally occur pair-wise are feminine in Norwegian. (ENGER, 2009, p. 1290)

The set in (7) exemplifies *large rules* and can be seen as simple because the entire system is accounted for with a few general rules. On the other hand, the set in (8) illustrates instances of *small rules*, which are small in scope, exception-ridden and conflicting. A language that employs *small rules* in order to organize its gender system needs a larger number of them in order to account for each and every noun. Therefore, languages with this type of assignment rule are considered complex.

The third dimension in which gender languages can be more or less complex is the formal expression of gender in the morphology of the language. Typically, the gender of a noun is opaque on the noun itself, as seen in English, but it is expressed via agreement on other words, such as the article, the adjective, the predicate, the participle and various pronouns.[2]

In some language, agreement is so ubiquitous that almost every word in the sentence carries a gender marking. In what follows, (9) is an example from Chichewa, a Bantu language from the Niger-Congo family, spoken in East-Africa (MCHOMBO, 2004, p. 87). Numbers 7, 1 and 9 indicate noun classes.

(9) Ichi   ndi chi-tsílu chi-méné kalulú a-na-chí-lémbélá   kálata
      7.DEF.SG be 7-fool   7-REL   1.hare 1-PST-7-write.to/for 9.letter
      'This is the fool that the hare wrote a letter to/for'

On the other hand, there are languages with extremely sparse expressions of gender. The best-known example is English, where gender is visible only on personal and possessive pronouns.

---

[2]Corbett (1991, p. 113) provides examples of more unusual targets, such as adpositions and complementizers.

3   Conclusion

In this squib, we tried to deal with the nature of gender complexity and how it can be conceived throughout the linguistic literature. In general, emphasis is placed on the empirical difficulties that need to be considered in assessing the nature of gender as a category in grammar. The second section discussed the morphosyntactic manifestation of gender as a (complex) feature, focusing on the difficulty of defining it as a traditional feature. Summarizing, agreement may be restricted in several ways. A common occurrence is that it interacts with other features by only surfacing in certain values of this feature (e.g., gender is only expressed in the singular in German, only in the third person in Serbo-Croatian, and only in the past tense in Russian). A second restricting factor is syncretism. Many agreement targets do not have dedicated forms for each paradigm cell. This reduces the likelihood that markers provide unambiguous gender information (see Romanian in Table 1 and Table 2). A final complicating factor is the form of the agreement markers themselves. This squib meant to review the discussion in the literature on the complexity of gender and to illustrate how complex this subject is.

References

AUDRING, J. Gender assignment and gender agreement: evidence from pronominal gender languages. *Morphology*, v. 18, 2009, p. 93-116.

AUDRING, J. Gender as a complex feature. *Language Sciences*. Special issue: Exploring grammatical gender, v. 43, 2014, p. 5-17.

CORBETT, G. *Gender*. Cambridge: Cambridge University Press, 1991.

CORBETT, G. *Agreement*. Cambridge: Cambridge University Press, 2006.

CORBETT, G. Number of Genders. In: DRYER, M.S.; HASPELMATH, M. (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013. Available at: http://wals.info/chapter/30. Accessed April 4, 2016.

CURZAN, A. *Gender Shifts in the History of English*. Cambridge: Cambridge University Press, 2003.

CYRINO, J. P. L.; ARMELIN, P. R. G.; MINUSSI, R. D. On the encyclopedic knowledge of gender. In: AUDRING, J.; KOUTSOUKOS, N.; MASINI, F.; RAFFAELLI, I. *Morphology and Semantics. MMM9 On-Line*. v. 1, 2013, p. 77-87.

DUKE, J. Gender reduction and loss in Germanic: the Scandinavian, Dutch and Afrikaans case studies. In: NÜBLING, D.; DAMMEL, A.; KÜRSCHNER, S. (eds.). *Kontrastive germanistische Linguistik*. Vol. II. Hildesheim: Georg Olms, 2010, p. 643-672.

ENGER, H.-O. The role of core and non-core semantic rules in gender assignment. *Lingua*, v. 119, 2009, p. 1281-1299.

HELLINGER, M.; BUßMANN, H. *Gender Across Languages*: The Linguistic Representation of Women and Men. Vol. 3. Amsterdam: John Benjamins, 2001.

HENGEVELD, K. Introduction: transparency in functional discourse grammar. *Linguistics in Amsterdam*, v. 4, n. 2, p. 1-22, 2011.

KRAMER, R. *Definite Markers, Phi Features and Agreement*: a Morphosyntactic Investigation of the Amharic DP. Doctoral Dissertation, University of California, Santa Cruz, 2009.

KRAMER, R. *The Morphosyntax of Gender*. Oxford: Oxford University Press, 2015.

MCHOMBO, S. *The Syntax of Chichewa*. Cambridge: Cambridge University Press, 2004.

MILLS, A. *The Acquisition of Gender*: A Study of English and German. Berlin: Springer, 1986.

MOORE, S. Grammatical and natural gender in Middle English. *PMLA*, v. 36, n.1, 1921, p. 79-103.

RESCHER, N. *Complexity*. A Philosophical Overview. New Brunswick and London: Transaction Publishers, 1998.

SIEMUND, P. *Pronominal Gender in English*: A Study of English Varieties from a Cross-linguistic Perspective. London: Routledge, 2008.

TRUDGILL, P. Language contact and the function of linguistic gender. *Poznan Studies in Contemporary Linguistics*, v. 35, 1999, p. 133-152.