Semantic versus lexical gender

Synchronic and diachronic variation in Germanic gender agreement

Kraaikamp, M.

Publication date
2017
Document Version
Other version
License
Other

Citation for published version (APA):

Chapter I
Introduction

Bij politici, sporters, presentatoren, op televisie bij DWDD en Pauw en op de radio hoor je elke dag: "het meisje ... die", een jongetje ... die" of "mijn paard ....die". Zo verwatert [...] kennelijk toch het Standaardnederlands.

'With politicians, athletes, hosts, on television in DWDD and Pauw and on the radio, every day, you hear: "the girl ... she", "a boy ... he" or "my horse ... he". Thus, Standard Dutch [...] apparently deteriorates.'

(Excerpt from a reader's e-mail to Dutch newspaper *De Telegraaf*, 25 January 2016, http://www.telegraaf.nl/watuzegt/25076464/HetisHETmeisjeDAT.html)

This dissertation is about gender agreement. The term 'gender' derives from the Latin word *genus*, meaning 'sort' or 'category'. It usually refers to the categories of male and female in the non-linguistic context, but in the present linguistic context, it is used in its broader sense. The nominal classes in language referred to as genders are not always based on a classification in terms of sex, but can also be based on, for instance, animacy or the noun's form. A good starting point to explain linguistic gender is Hockett's (1958: 231) definition: "Genders are classes of nouns reflected in the behaviour of associated words". This behaviour of associated words is called gender agreement. Corbett's (1991) extensive reference work on gender shows that gender systems are found in many different languages of the world, varying in the number of genders used, their basis of classification, and the kind and number of elements that show gender agreement. Many of the Indo-European languages have a gender system, inherited from Proto-Indo-European, which generally involves two or three nominal genders and agreement on determiners, adjectives and pronouns.

The dissertation focuses on pronominal gender agreement in Dutch, a Germanic language with two nominal genders in its standard variety spoken in The Netherlands. The outline of this chapter is as follows. Section 1 introduces the concepts of gender assignment and gender agreement. Section 2 discusses gender

agreement in pronouns and the agreement variation that pronouns can display. Section 3 introduces the gender system of Dutch and the agreement phenomenon that is the subject of this study. Section 4 presents the aim and research questions of this dissertation and gives an overview of the outline of this book.

1. Gender assignment and gender agreement

In languages with nominal genders, each noun belongs to a certain gender. In comparison with other grammatical categories, nominal gender is a remarkable one, because, although it forms an integral part of many languages around the world, it does not seem to serve a grammatical function in most languages. The gender of a noun is invariant in most gender systems, that is, speakers cannot choose the gender of a noun, in the same way that they can choose, for instance, the singular or plural number of a noun. Nominal gender does not vary from one context to the next and does not modify the meaning of the noun. As such, nominal gender does not behave as most grammatical categories, which tend to have a choice of meaningful settings (Leiss 2000). While the purpose of nominal gender is obscure, it can be one of the most cumbersome aspects of a language to learn. This is the case for Dutch in particular, where the gender of nouns largely has to be learnt word by word, which takes a relatively long time even for native speakers and is a feat that is often never fully accomplished by second language learners (Blom, Polišenská & Weerman 2008).

The attribution of a particular gender to a noun is called gender assignment. Gender assignment can be based on different properties of the noun, viz. its semantic properties or its form, that is, the morphological or phonological properties of the noun, and often a combination of form and meaning is involved (Corbett 1991). An example of a language with a purely semantic gender assignment system is Diyari, an Australian Aboriginal language with two genders, where nouns denoting female humans and female animals belong to one gender, and all other nouns belong to the other gender (Corbett 1991: 11). Assignment systems that are purely based on form, without any semantic factors involved, do not exist, as all gender systems show at least some semantic assignment rules (Corbett 1991, Corbett & Fraser 2000). An example of a language with a predominantly formal gender assignment system is Qafar, an East Cushitic language with two genders, where

nouns that end in an accented vowel belong to one gender and all other nouns, those ending in a consonant or unaccented vowel, belong to the other gender (Corbett 1991: 51). In some languages, gender assignment is not systematically based on formal or semantic criteria. This can be said for Dutch, which shows some regularities in gender assignment, such as that nouns referring to countries are usually neuter gender or nouns ending in -*de* or –*te* are usually common gender, but these regularities cover only a very limited portion of the lexicon. Gender assignment to the majority of Dutch nouns does not follow any formal or semantic criteria, but is arbitrary (Donaldson 1987, Haeseryn, Romijn, Geerts, De Rooij & Van den Toorn 1997, Booij 2002).

Following Hockett's (1958) definition of gender, gender systems are defined by the existence of gender agreement. Agreement in general can be described as the "systematic covariance between a semantic or formal property of one element and a formal property of another" (Steele 1978: 610). The first element controls the agreement and determines the form of the latter element, the agreement target (Corbett 2006: 4). In the case of gender agreement, the agreement controller is typically a noun and possible agreement targets include determiners, numerals, adjectives, verbs and pronouns (Corbett 1991: 105-112). The agreement targets are able to take different forms and can show gender agreement with the controller in this way. An example of gender agreement in Italian is shown in (1) below, where the demonstrative determiner, predicative adjective and pronoun all show agreement with the masculine gender of the noun *libro* 'book'.

(1) Quest-o libro è molt-o stran-o. Vorrei   legger-lo.
     DEM-M book is very-M strange-M would.like.1SG to.read-3SG.M
     'This book is very strange. I would like to read it.'

Although nominal gender assignment and gender agreement are concepts that are distinguished from each other, the two are intertwined, as a noun's gender is ultimately defined by the agreement it receives. The noun *libro* in example (1) is considered a masculine noun, because it is combined with masculine forms. The relevance of gender agreement to nominal gender assignment becomes particularly clear in cases where the agreement that a noun receives is different from the gender

that is typically associated with the noun's form. The noun *mano* 'hand' in Italian is an example of this. Although nouns ending in *–o* are usually masculine in Italian, *mano* combines with feminine forms, such as the determiner *la*, and it is considered a feminine, not a masculine, noun. In Dutch, most nouns do not even have a gender-specific form that could indicate their gender. The interconnectedness of gender assignment and gender agreement is illustrated in Dutch by the fact that speakers informally refer to common and neuter nouns in Dutch as *de-woorden* '*de*[DET.C] words' and *het-woorden* '*het*[DET.N] words', distinguishing the nouns on the basis of the gender form of the definite article they take.

2. Pronominal gender agreement and agreement variation

This thesis focuses on gender agreement in pronouns. Pronouns are exceptional agreement targets for two reasons. One is that, unlike other agreement targets, they can be widely separated from the noun that controls the agreement. Pronouns do not have to occur in the same sentence as the antecedent noun and can even be separated from it by a speaker turn, as in the dialogue in (2):

(2)   Did you like the book I gave you for your birthday? – Yes, I loved it. I just finished it yesterday.

Another aspect of pronouns that makes them exceptional as agreement targets is that they are coreferential with the agreement-controlling noun, that is, the pronoun and the noun share a real-world referent. The pronominal references in examples (1) and (2) above are called 'anaphoric reference', which means that there is a linguistic antecedent to which the pronoun refers 'back', viz. the noun *libro* 'book' in (1) and *book* in (2). This is different with deictic reference, where there is no linguistic antecedent and the pronoun refers to the real-world referent directly, for example when a speaker points and says 'look at him'. Although the criterion of the presence or absence of a linguistic antecedent makes a clear distinction between the types of pronominal reference, they are not entirely distinct. Even in the presence of an antecedent noun, a pronoun does not necessarily refer only to this noun, but, as in deictic reference, likely refers to the referent directly as well (Lyons 1977: 646-677). Lyons (1977: 660) describes anaphoric reference as follows: "[w]e

4

will not say that a pronoun refers to its antecedent but rather that it refers to the referent of the antecedent expression with which it is correlated". While the pronoun and the antecedent are connected in the discourse, they still each on their own refer to the referent in the real world.

Thus, anaphoric pronouns have a connection with both the antecedent noun and their real-world referent. It may be this dual connection that makes pronouns particularly prone to displaying agreement variation. An example of pronominal agreement variation in German is shown in (3a) and (3b):

(3)   a.   Dieses **Mädchen** ist sehr sympathish. Ich möchte   **es**   gerne
           DEM.N girl(N)   is   very sympathetic I   would.like 3SG.N gladly

           kennen-lernen.
           to.know-learn

           'This girl is very sympathetic. I would like to get to know her.'

      b.   Dieses **Mädchen** ist sehr sympathish. Ich möchte   **sie**   gerne
           DEM.N girl(N)   is   very sympathetic I   would.like 3SG.F gladly

           kennen-lernen.
           to.know-learn

           'This girl is very sympathetic. I would like to get to know her.'

The pronoun in the German example above can be either neuter, as in (3a), or it can be feminine, as in (3b). The pronominal agreement in (3a) is in accordance with the neuter gender of the antecedent noun *Mädchen* 'girl'. Different terms are used in the literature for this canonical type of gender agreement. Throughout this thesis it is referred to as 'lexical gender agreement', following Dahl (2000). It is also known as 'grammatical gender agreement' (e.g. Baron 1971) or 'syntactic gender agreement' (e.g. Corbett 1991, Audring 2009). However, these terms are not always clearly or appropriately defined. 'Grammatical gender agreement' can also refer to gender

agreement in general, just as 'grammatical gender' is often used for linguistic gender in general (for instance, Corbett 1991, Comrie 1999). Corbett (1991: 226) defines 'syntactic agreement' as "agreement consistent with form, that is, agreement consistent with the gender as it would be assigned by morphological or phonological assignment rules". However, this definition is a little too restricted, especially to describe gender agreement in a language like Dutch, as not all nouns have a gender-specific form. Dahl (2000) also points out this shortcoming of Corbett's terminology and proposes that what Corbett refers to as 'syntactic agreement' is agreement with 'lexical gender', that is, agreement that is based on the lexically stored gender of the noun.

The feminine agreement in (3b) is in not in accordance with the gender of the antecedent noun, but instead it is in accordance with the gender that is associated with the referent. This kind of agreement is known as 'semantic gender agreement'. This term is commonly used in the literature and therefore it is used throughout this thesis as well. However, it would perhaps be more accurately called agreement based on 'referential gender', following Dahl (2000). The term 'semantic gender agreement' is in line with Corbett's (1991) description of it as agreement with the semantic properties of the noun, which, in the case of (3b), would be the semantic feature [female] of the noun *Mädchen* 'girl'. However, rather than being based on the semantic properties of the noun, this kind of agreement is more likely based on the properties of the referent, that is, the sex of the actual girl in example (3). This becomes apparent with nouns whose meaning does not specify the sex of the referent. The English noun *doctor* is an example of this. Pronominal agreement with either *he* or s*he* with this noun varies from one context to the next, depending on whether the doctor in question is a man or a woman. This indicates that the pronominal agreement is not based on the semantic properties of the noun *doctor*, which does not include a sex specification, but on the properties of the actual referent in context (Dahl 2000). A consequence of this analysis is that with semantic gender agreement, it is not the noun that is the agreement controller, but the referent is. In this view, variation between lexical and semantic gender agreement, as in (3a) and (3b), is variation between agreement that is based on the gender assigned to the noun and agreement that is based on the gender associated with the referent.

3. Semantic gender agreement in Dutch

This thesis studies pronominal gender agreement in Dutch. Standard Dutch spoken in The Netherlands (subsequently referred to as 'Dutch') has two nominal genders, common and neuter. Common gender is a conflation of former masculine and feminine gender. These two genders conflated around the seventeenth century as a result of deflection in the noun phrase (Geerts 1966: 192-210, Schönfeld & Van Loey 1970: 119-120). Several Eastern and Southern Dutch dialects, particularly Flemish dialects, still distinguish the original three nominal genders to varying extents, but the distinction between masculine and feminine nouns is no longer made in Standard Dutch spoken in The Netherlands. Gender agreement in Dutch is shown on determiners, such as the definite article (*de* versus *het*), attributive adjectives and pronouns. In the pronominal domain, the two genders are expressed on the relative pronoun (common *die* versus neuter *dat* 'that') and on the demonstrative pronouns (common *die, deze* versus neuter *dat, dit* 'that, this'). The personal pronoun distinguishes masculine (*hij*, *hem* 'he, him'), feminine (*zij*, *haar* 'she, her') and neuter (*het* 'it') gender.

As in the German example (3) above, pronouns in Dutch show variation between lexical gender agreement and semantic gender agreement. Interestingly, semantic agreement in Dutch not only occurs with animate referents, but it occurs with inanimate referents as well. This has been observed by, among others, Van Haeringen (1936, 1951), Fletcher (1987), Siemund (2002) and Audring (2006, 2009). Audring (2009) demonstrates and analyses the phenomenon in great detail on the basis of spoken language data from the *Corpus Gesproken Nederlands* (*CGN*, 'Corpus of Spoken Dutch'). The semantic agreement with inanimates appears to be based on the degree of individuation of the referent. Common and masculine pronouns are used with referents that have a high degree of individuation, that is, referents with a clearly bounded shape, such as concrete objects, while neuter pronouns are used with referents that have a low degree of individuation, that is, referents with unclear boundaries, such as materials and liquids. Examples of this semantic agreement from the *CGN* are shown in (4) and (5) below (from Audring 2006: 95-99). In (4) a masculine pronoun is used with a neuter gender noun denoting a concrete object, *boek* 'book', while in (5) a neuter pronoun is used with a common gender noun denoting a mass, *olijfolie* 'olive oil'.

(4) Moet je nog wat informatie over dat **boek** hebben?
  need you more some information about DEM.N book(N) have

  Dan moet 'k **'m** nog niet gaan inleveren.
  then should I 3SG.M yet not go return

  'Do you need some more information about that book? Then I shouldn't return
  it yet.'

(5) 't zit toch ook bij **olijfolie** wel een beetje in
  it is in.fact also with olive.oil(C) PRT a bit about

  hoe **'t** geconserveerd wordt.
  how 3SG.N preserved is

  'In fact also with olive oil, it matters how it is preserved.'

This semantic gender agreement exists beside lexical gender agreement in present-
day Dutch. For example, the pronoun in (4) could alternatively be neuter, as in (4'),
agreeing with the neuter gender of the noun *boek*:

(4') Moet je nog wat informatie over dat **boek** hebben?
  need you more some information about DEM.N book(N) have

  Dan moet 'k **'t** nog niet gaan inleveren.
  then should I 3SG.N yet not go return

  'Do you need some more information about that book? Then I shouldn't return
  it yet.'

This thesis was inspired by Audring's work on pronominal gender agreement
in Dutch and investigates the origin of the observed semantic agreement based on

individuation, when this kind of agreement has developed and what factors could be involved in its surfacing. Audring (2006, 2009) herself proposes that the semantic agreement has developed in response to the change from a three-gender system to a two-gender system in Dutch, due to the conflation of masculine and feminine nominal gender. The resulting mismatch between nominal gender and pronominal gender may have led to a reinterpretation of the pronouns in semantic terms. This dissertation explores an alternative scenario. It investigates the possibility that agreement based on individuation already existed before this change, and that the semantic agreement that is now observed in pronouns reflects a long existing semantic interpretation of the genders.

4. Aim and outline

The aim of this dissertation is to gain insight into the origin of semantic gender agreement based on individuation in Dutch, when it has developed and what factors could be involved in its surfacing. This thesis addresses the following research questions that are each intended to shed light on this issue:

(i)   How deeply rooted is the association of common and masculine gender with the meaning of high individuation and neuter gender with the meaning of low individuation in Dutch? Is this semantic association restricted to pronominal gender or can it be found in nominal gender as well? Could the association go back to a semantic origin of the gender system?

(ii)   Does semantic agreement based on individuation also exist in Germanic varieties that still distinguish the original three nominal genders?

(iii) Did semantic agreement based on individuation develop in Dutch after the change from a three-gender system to a two-gender system or did it exist already before this change?

(iv)   Is there a relation between the extent to which lexical gender is marked in the noun phrase and the ratio of semantic gender agreement in pronouns?

The core of this dissertation consists of four chapters that each address one of the research questions presented above. These chapters were written as separate research papers that have been published in or submitted to different linguistics journals. This means that the chapters can be read independently and, although they each address a separate research question, there is some overlap between them.

Chapter II addresses the first research question, or set of questions. In this chapter it is argued that individuation plays a role in the Dutch gender system as a whole. The association between the genders and different degrees of individuation not only exists in pronouns but can be found in nominal gender in Dutch as well. The semantic agreement found in pronouns relates to an existing semantic distinction between the genders, which is also found in the gender systems of other Germanic, and Romance, varieties and possibly dates back to Proto-Indo-European.

Chapter III addresses the second research question and investigates whether semantic agreement based on individuation also exists in German, a Germanic variety that still distinguishes three nominal genders. The chapter presents an pronoun elicitation experiment that compares pronominal agreement in Dutch and in German. The results of the experiment show that semantic agreement based on individuation exists in both languages, but in considerably different ratios.

Chapter IV addresses the third research question of when semantic agreement based on individuation developed in Dutch pronouns, particularly whether or not it existed before the conflation of masculine and feminine nominal gender. This chapter presents a corpus study of pronominal agreement in Middle Dutch, where the original three-gender system was still in place. The results show that agreement based on individuation already existed beside lexical gender agreement in Middle Dutch, although its frequency seems to be lower than in present-day Dutch.

Chapter V addresses the final, fourth research question. It investigates whether the visibility of lexical gender in the noun phrase could influence the ratio of lexical to semantic gender agreement in pronouns. The experiment presented in this chapter tests whether the presence or absence of explicit lexical gender marking on the antecedent affects the choice between lexical and semantic gender agreement in pronouns. The results of the experiment show that the absence of lexical gender marking increases the likelihood of semantic agreement.

Chapter VI summarizes and discusses the findings of this dissertation, addresses remaining questions and provides suggestions for future research.