



## UvA-DARE (Digital Academic Repository)

### Semantic versus lexical gender

*Synchronic and diachronic variation in Germanic gender agreement*

Kraaikamp, M.

#### Publication date

2017

#### Document Version

Other version

#### License

Other

[Link to publication](#)

#### Citation for published version (APA):

Kraaikamp, M. (2017). *Semantic versus lexical gender: Synchronic and diachronic variation in Germanic gender agreement*. [Thesis, fully internal, Universiteit van Amsterdam]. LOT.

#### General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

## Summary

### Semantic versus lexical gender: synchronic and diachronic variation in Germanic gender agreement

Most Germanic languages, including Dutch, have a gender system in which each noun belongs to a certain gender. These nominal genders are reflected by gender agreement on associated words, such as determiners, adjectives and pronouns. However, pronouns do not always show gender agreement with their antecedent noun. A familiar example of this is when a masculine or feminine pronoun is used with a neuter noun that refers to a human being. For example, in Dutch the masculine pronoun *hij/hem* ‘he/him’ can be used with the neuter noun *jongetje* ‘boy’ or the feminine pronoun *zij/haar* ‘she/her’ with the neuter noun *meisje* ‘girl’. In these cases the pronoun shows what is called ‘semantic gender agreement’, agreement that is based on the properties of the referent rather than the lexical gender of the noun. Another, perhaps less familiar, kind of semantic gender agreement that occurs in Dutch involves inanimate referents. Examples of this are shown in (1) and (2) below, where in (1) a masculine pronoun (*hem*) is used with a neuter noun, *boek* ‘book’, and in (2) a neuter pronoun (*het*) is used with a common gender noun, *gember* ‘ginger’:

- (1) *Ik heb m'n boek nog niet uit, maar ik moet hem terugbrengen naar de bibliotheek.*

‘I have not finished my book yet, but I have to return it to the library.’

- (2) *Je kunt verse gember bijna overal krijgen, maar ik vind het vaak te duur.*

‘You can get fresh ginger almost everywhere, but I think it is often too expensive.’

The agreement pattern that these examples illustrate involves the degree of individuation of the referent. Masculine and common gender pronouns, on the one hand, tend to be used with referents that have a high degree of individuation, that is, referents that have a clearly bounded shape, such as the object *boek* ‘book’ in (1).

## SUMMARY

Neuter gender pronouns, on the other hand, tend to be used with referents that have a low degree of individuation, that is, referents with unclear boundaries, such as the mass *gember* ‘ginger’ in (2).

The aim of this dissertation is to investigate the origin of this semantic agreement pattern in Dutch, when it has developed and what factors could be involved in its surfacing. This work consists of four separate studies that are intended to shed light on the roots of the phenomenon. The first study explores the semantic underpinnings of the Dutch gender system. The second study is an experimental study that compares the agreement behaviour in Dutch with that in German. The third study is a historical corpus study of pronominal gender agreement in Middle Dutch. The fourth and final study is an experimental study that investigates the effect of adnominal gender marking on pronominal agreement variation.

The first study shows that the association of common and neuter gender with a high and a low degree of individuation, respectively, not only exists in pronouns, but can be found in nominal gender as well. Although nominal gender assignment is largely semantically arbitrary in Dutch, a semantic pattern surfaces in certain cases. This occurs in particular when the gender of the noun is not fixed but variable. The association between neuter gender and a low degree of individuation is also visible when the antecedent is not a noun, but, for instance, a predicate. These observations indicate that the semantic interpretation of the genders is not an innovation in Dutch pronouns, but something that has long existed. There are in fact indications that the semantic interpretation of the genders goes back to their roots in Proto-Indo-European. In line with this, semantic agreement based on individuation has been found not only in Dutch, but also in other Germanic varieties, including dialects of English, Frisian and Danish, and in Romance dialects. The Indo-European gender system possibly had a semantic basis, which has become disrupted in the nominal domain ever since nominal gender became an invariable, lexically stored feature of nouns. The lexical gender of nouns is consequently no longer always in line with their meanings and as a result of this, a conflict can arise between the lexical gender of nouns and the semantic gender associated with their referents.

The second, experimental, study is aimed at investigating the possible relation between pronominal agreement based on individuation and a change that took place in the Dutch nominal gender system: the change from three to two nominal genders

as a result of the conflation of masculine and feminine gender into one common gender. Agreement based on individuation has so far only been attested in Germanic varieties in which masculine and feminine gender have conflated, as in Dutch, or that lost lexical gender altogether. The second study investigates whether agreement based on individuation exist also in German, a Germanic variety that still distinguishes the original three nominal genders, and how pronominal agreement in German compares to that in Dutch. This study involves pronoun elicitation experiments in Dutch and in German, in which native speakers of each language are prompted to refer to particular referents that differ in their degree of individuation. The results show that both speakers of German and speakers of Dutch are inclined to show semantic neuter agreement with masses and abstract referents, and semantic masculine agreement with animal referents. However, semantic masculine agreement with object referents is shown only by speakers of Dutch and also the frequency of semantic agreement is overall much higher in Dutch than in German. The findings of this study suggest that semantic agreement based on individuation is a shared Germanic feature and that the difference between Dutch and German with respect to this kind of agreement is mainly one of degree.

The third, corpus, study investigates the diachrony of agreement based on individuation in Dutch and focuses on the question whether or not this kind of agreement already existed in Dutch before the distinction between masculine and feminine nominal gender was lost. This study is a corpus study of pronominal agreement in Middle Dutch recipe books from the early 16<sup>th</sup> century. The results show that semantic agreement based on individuation existed already in Middle Dutch, in particular, semantic neuter agreement with mass referents. Semantic agreement with object referents is not attested in the corpus. It appears that in Middle Dutch, semantic agreement occurs only with referents that have either a very high degree of individuation, that is, animate referents, or that have an extremely low degree of individuation, that is, masses. A comparison of the frequency of semantic agreement in Middle Dutch to that in present-day Dutch suggests that semantic agreement has generally increased over time.

The final, experimental, study investigates a factor that could play a role in the diachronic increase of semantic gender agreement, as well as the synchronic variation in its frequency between languages such as Dutch and German: the

## SUMMARY

visibility of lexical gender in the noun phrase. Dutch has changed over time in this respect and now differs from German, as deflection has led to a loss of lexical gender marking on several elements inside the noun phrase, such as the indefinite article and possessive determiners. The fourth, final, study investigates whether there is a relation between lexical gender marking in the noun phrase and the choice between semantic and lexical gender agreement in pronouns. This study involves a pronoun elicitation experiment with speakers of Dutch in which the test sentences vary in one crucial aspect: the antecedent noun is accompanied by either a determiner carrying explicit gender marking or a determiner without gender marking. The results show that the likelihood of semantic agreement is higher when there is no gender marking on the antecedent. This finding reveals that the absence of lexical gender marking in the noun phrase makes semantic agreement surface more easily in pronouns. This effect can explain both the synchronic variation between semantic and lexical gender agreement and the shift in the frequency of semantic agreement in Dutch over time.

The results of the four studies that are presented in this dissertation indicate that the semantic gender agreement that is observed in Dutch pronouns today relates to an existing semantic interpretation of the genders, which possibly goes back to the semantic roots of the gender system. Pronominal agreement based on individuation existed already in Middle Dutch and it surfaces in varying degrees in other Germanic varieties, including German. The rigid nature of the lexical gender system, that is, the fact that the lexical gender of nouns is largely fixed and invariable, creates a conflict between the lexical gender of nouns and the semantic gender that is associated with their referents. The extent to which semantic gender agreement can surface in pronouns, that is, the ratio of semantic to lexical gender agreement in pronouns, appears to depend on the visibility of lexical gender in the noun phrase. As this aspect of the lexical gender system varies both synchronically and diachronically, it can explain both the variation in the ratio of semantic to lexical gender agreement between languages and the increase of semantic gender agreement in Dutch over time.

## Samenvatting

### Semantisch versus lexicaal geslacht: synchrone en diachrone variatie in Germaanse geslachtscongruentie

De meeste Germaanse talen, waaronder het Nederlands, hebben een geslachtssysteem waarin elk nomen tot een bepaald geslacht behoort. Deze lexicale geslachten zijn terug te zien in congruentie op andere elementen, zoals lidwoorden, bijvoeglijke naamwoorden en voornaamwoorden. Voornaamwoorden vertonen echter niet altijd congruentie met het woordgeslacht van het antecedent. Een bekend voorbeeld hiervan is het gebruik van een mannelijk of vrouwelijk voornaamwoord bij een onzijdig zelfstandig naamwoord dat verwijst naar een persoon. Zo kan in het Nederlands een mannelijk voornaamwoord (*hij/hem*) gebruikt worden bij een onzijdig nomen zoals *jongetje* of een vrouwelijk voornaamwoord (*zij/haar*) bij een onzijdige nomen zoals *meisje*. In dit geval vertoont het voornaamwoord zogenaamde ‘semantische congruentie’, congruentie die gebaseerd is op de eigenschappen van de referent in plaats van het lexicale geslacht van het zelfstandig naamwoord. Een ander, wellicht minder bekend, soort semantische congruentie in het Nederlands treedt op bij onbezielde referenten. De voorbeelden (1) en (2) hieronder laten dit fenomeen zien. In (1) wordt een mannelijk voornaamwoord, *hem*, gebruikt bij het onzijdige nomen *boek* en in (2) wordt een onzijdig voornaamwoord, *het*, gebruikt bij het commune nomen *gember*.

(1) *Ik heb m'n boek nog niet uit, maar ik moet hem terugbrengen naar de bibliotheek.*

(2) *Je kunt verse gember bijna overal krijgen, maar ik vind het vaak te duur.*

Deze voorbeelden illustreren een congruentiepatroon dat te maken heeft met de individueringsgraad van de referent. Referenten met een hoge mate van individuering hebben een duidelijk begrensde vorm, zoals personen, dieren en voorwerpen. Het *boek* in (1) is een voorbeeld hiervan. Referenten met een lage individueringsgraad hebben daarentegen juist een vage begrenzing, zoals abstracte

## SAMENVATTING

zaken en stoffen. Een voorbeeld hiervan is de *gember* in (2). In het Nederlands bestaat de tendens om met mannelijke en commune voornaamwoorden te verwijzen naar referenten die een hoge individueringsgraad bezitten en met onzijdige voornaamwoorden naar referenten die een lage individueringsgraad bezitten.

Dit proefschrift heeft als doel uit te zoeken wat de oorsprong van dit semantische congruentiepatroon in het Nederlands is, wanneer het zich heeft ontwikkeld en welke factoren betrokken zouden kunnen zijn bij het verschijnen ervan. Dit werk bestaat uit vier verschillende onderzoeken die bedoeld zijn om de wortels van dit fenomeen te ontdekken. Het eerste onderzoek verkent in hoeverre semantiek nog meer een rol speelt in het Nederlandse geslachtssysteem en of het systeem mogelijk een semantische oorsprong heeft. Het tweede onderzoek is een experimenteel onderzoek waarin het verwijsgedrag in het Nederlands wordt vergeleken met dat in het Duits. Het derde onderzoek is een historisch corpusonderzoek naar het voornaamwoordgebruik in het Middelnederlands. Het vierde, laatste, onderzoek betreft een experiment waarin het effect wordt onderzocht van geslachtsmarkering rond het nomen op de congruentie in voornaamwoorden.

Het resultaat van het eerste onderzoek laat zien dat de connectie tussen commuun geslacht en een hoge individueringsgraad, enerzijds, en onzijdig geslacht en een lage individueringsgraad, anderzijds, niet alleen bestaat bij voornaamwoorden, maar ook in het nominale domein. Hoewel de geslachtstoekenning aan zelfstandige naamwoorden grotendeels arbitrair is in het Nederlands, is er in sommige gevallen toch een semantisch patroon zichtbaar. Dit is met name het geval wanneer het geslacht van een zelfstandig naamwoord niet vastligt, maar kan variëren. Daarnaast is de connectie tussen onzijdig geslacht en lage individuering ook te zien in gevallen waarin het antecedent van een voornaamwoord geen zelfstandig naamwoord is maar, bijvoorbeeld, een predicatuur. De bevindingen duiden erop dat de semantische interpretatie van de geslachten geen innovatie is van de voornaamwoorden, maar al langer onderdeel is van het Nederlandse geslachtssysteem. Er zijn zelfs aanwijzingen dat de semantische interpretatie van de geslachten teruggaat op hun oorsprong in het Proto-Indo-Europees. Wat hiermee in overeenstemming is, is dat semantische congruentie op basis van individuering niet alleen in het Nederlands wordt gevonden, maar ook in andere Germaanse talen, waaronder dialecten van het Engels, Fries en Deens, en in

Romaanse dialecten. De mogelijkheid bestaat dat het Proto-Indo-Europese geslachtssysteem een semantische basis had die verstoord is geraakt doordat nominaal geslacht niet langer een variabele waarde is, maar veelal een onveranderlijk, lexicaal opgeslagen kenmerk van zelfstandige naamwoorden. Dit maakt dat het geslacht van nomina niet altijd meer in overeenstemming is met hetgeen waarnaar ze verwijzen. Als gevolg hiervan kan er een conflict ontstaan tussen het lexicale geslacht van zelfstandige naamwoorden en het semantische geslacht dat bij hun referent past.

Het tweede, experimentele, onderzoek richt zich op de mogelijke relatie tussen het optreden van semantische congruentie op basis van individuering in voornaamwoorden en een verandering in nominale geslachtssysteem van Nederlands: de verandering van drie naar twee nominale geslachten door het samenvallen van het mannelijke en vrouwelijke woordgeslacht tot één gemeen geslacht. Congruentie op basis van individuering is tot nu toe alleen gevonden in Germaanse talen waarin mannelijk en vrouwelijk geslacht, net als in het Nederlands, zijn samengevallen of waarin de woordgeslachten helemaal verdwenen zijn. Het tweede onderzoek bekijkt of semantische congruentie op basis van individuering ook plaatsvindt in het Duits, een Germaanse taal die nog wel drie woordgeslachten onderscheidt, en in hoeverre het verwijsgedrag in het Duits verschilt van dat in het Nederlands. Dit onderzoek betreft een Nederlandstalig en een Duitstalig voornaamwoord-eliciterings-experiment, waarin moedertaalsprekers van de twee talen worden uitgelokt om te verwijzen naar bepaalde referenten die van elkaar verschillen qua individueringsgraad. Uit dit onderzoek blijkt dat sprekers van het Duits, net als sprekers van het Nederlands, geneigd zijn onzijdige voornaamwoorden te gebruiken voor stoffen en abstracta, en mannelijke voornaamwoorden voor dieren. Het gebruik van mannelijke voornaamwoorden voor voorwerpen wordt echter alleen bij sprekers van het Nederlands gezien en ook is in het algemeen de frequentie waarmee semantisch gecongrueerd wordt veel hoger in het Nederlands dan in het Duits. De resultaten van dit onderzoek geven aan dat semantische congruentie op basis van individueringsgraad een gedeeld Germaans kenmerk is en dat het verschil tussen het Nederlands en het Duits met betrekking tot dit soort congruentie voornamelijk gradueel van aard is.

## SAMENVATTING

Het derde onderzoek betreft een corpusonderzoek naar de diachrone ontwikkeling van semantische congruentie in het Nederlands, gericht op de vraag of congruentie op basis van individuering al bestond in het Nederlands voordat het onderscheid tussen het mannelijke en vrouwelijke woordgeslacht verdwenen was. Dit onderzoek richt zich op het Middelnederlands en kijkt naar het gebruik van voornaamwoorden in Middelnederlandse receptenboeken uit het begin van de zestiende eeuw. De resultaten laten zien dat semantische congruentie op basis van individuering al in het Middelnederlands bestond, met name het gebruik van onzijdige voornaamwoorden voor stoffen. Het gebruik van mannelijke voornaamwoorden voor objecten wordt niet vertoond in het corpus. Semantische congruentie lijkt in het Middelnederlands alleen voor te komen bij referenten die ofwel heel hooggeïndividueerd zijn, zoals bezielde referenten, ofwel heel laaggeïndividueerd, zoals stoffen. Als de frequentie waarmee semantische congruentie voorkomt in het Middelnederlands wordt vergeleken met die in het huidige Nederlands, lijkt het erop dat semantische congruentie met de tijd is toegenomen.

Het laatste, experimentele, onderzoek, richt zich op een factor die een rol zou kunnen spelen bij zowel de diachrone toename van semantische congruentie als de synchrone variatie in de frequentie ervan tussen talen als het Nederlands en het Duits: de zichtbaarheid van lexicaal geslacht in het nominale domein. Het Nederlands is wat dit betreft veranderd en verschilt hierin nu van het Duits. Deflectieprocessen hebben in het Nederlands namelijk geleid tot een verlies van geslachtsmarkering op verscheidene elementen rond het nomen, zoals het onbepaalde lidwoord en de bezittelijke voornaamwoorden. In het vierde, laatste, onderzoek wordt gekeken of er een relatie bestaat tussen de markering van lexicaal geslacht in het nominale domein en de keuze tussen semantische en lexicale congruentie in het voornaamwoord. Dit onderzoek betreft een voornaamwoord-elicitering-experiment met sprekers van het Nederlands, waarbij de testzinnen van elkaar verschillen in één cruciaal opzicht: Het antecedent van het voornaamwoord bevat een element dat het geslacht van het nomen markeert of een element dat dit juist niet doet. De resultaten laten zien dat de kans op semantische congruentie hoger is als het antecedent geen geslachtsmarkering bevat. Deze bevinding laat zien dat de afwezigheid van geslachtsmarkering in het nominale domein semantische

congruentie makkelijker naar boven laat komen. Dit effect kan zowel de synchrone variatie tussen semantische en lexicale geslachtscongruentie verklaren als de toename van semantische congruentie in het Nederlands door de tijd heen.

De resultaten van de vier onderzoeken die in dit proefschrift beschreven worden geven aan dat de semantische congruentie die tegenwoordig bij Nederlandse voornaamwoorden gezien wordt geen geïsoleerd fenomeen is, maar verband houdt met een bestaande semantische interpretatie van de geslachten, die mogelijk teruggaat op de semantische wortels van het geslachtssysteem. Congruentie op basis van individuering bestond al in het Middelnederlands en verschijnt in wisselende mate in andere Germaanse talen, waaronder het Duits. De starre aard van het lexicale geslachtssysteem, dat wil zeggen, het feit dat het lexicale geslacht van zelfstandige naamwoorden grotendeels vastligt en niet kan variëren, schept een conflict tussen het lexicale geslacht van nomina en het semantische geslacht dat geassocieerd wordt met hun referent. De mate waarin semantische congruentie kan verschijnen in voornaamwoorden, oftewel de ratio van semantische congruentie ten opzicht van lexicale congruentie bij voornaamwoorden, blijkt afhankelijk te zijn van de zichtbaarheid van lexicaal geslacht in het nominale domein. Aangezien dit aspect van het lexicale geslachtssysteem zowel synchrone als diachrone variatie vertoont, kan dit zowel het verschil in het voorkomen van semantische congruentie tussen talen verklaren als de toename ervan in het Nederlands door de tijd heen.