Pronominal gender in Dutch
Apparent time change in lexical vs semantic agreement

Kristel Doreleijers[1,2] *Jeroen van Craenenbroeck*[1,3] *Marjo van Koppen*[1,4]

[1]Meertens Institute/LiME [2]Tilburg University/DCU [3]KU Leuven/CRISSP [4]Utrecht University/ILS

March 20, 2023

Abstract

This paper discusses pronominal gender agreement in Dutch. Based on a sentence completion task filled out by more than 10,000 speakers, we provide evidence that there is a shift from lexical to semantic gender agreement in Dutch, as previously suggested by Audring (2006), even in a formal register. Results of a correspondence and cluster analyses indicate that nouns with the same degree of individuation group together. Furthermore, the analyses reveal three distinct speaker groups that follow a specific gender agreement pattern. Younger speakers are more semantically oriented than older speakers, who are more lexically oriented, which points to apparent time language change.

1 Introduction

This paper[1] deals with pronominal gender in Dutch, i.e., the use of the personal pronouns *hij* 'he', *zij* 'she', and *het* 'it' to refer back to a noun phrase in the preceding discourse (Haeseryn et al. 2019:§5.2.5). The pronominal gender system in Standard Dutch is a grammatical (lexical) system that establishes agreement between the lexical gender of a noun and its referent (De Paepe and De Vogelaer 2008:4). For example, the sentence in (1) can have three different meanings depending on which pronoun is used: *hij* 'he' refers to *de secretaris* 'the secretary', *zij* 'she' to *de koningin* 'the queen', and *het* 'it' to *het boek* 'the novel'.

(1) De film naar aanleiding van het boek over de secretaris van de koningin is gemaakt toen
the movie on occasion of the book about the secretary of the queen is made when
hij/zij/het vijftig jaar oud was.
he/she/it fifty year old was
'The movie based on the novel about the queen's secretary was made when he/she/it was fifty years old.'

Speakers may struggle with their pronoun selection if they are already unsure about the lexical gender of the noun. Not only do they have to know the grammar rules, they also have to store the lexical gender of individual nouns in memory (De Paepe and De Vogelaer (2008:5); see also Brouwer et al. (2017) for a processing study). In Dutch, there are very few overt indications of the lexical gender of a noun. Most of the time it cannot be read off the linguistic form directly, except for words ending in a diminutive suffix (which are always of neuter gender, e.g., *de koek* 'the cake' - *het koekje* 'the cookie') or in derivational suffixes that are typical of feminine lexical gender such as *-heid*, *-ie*, *-te*, or *-ing* (Haeseryn et al. 2019:§3.3.3). In addition, nouns that refer to entities with a biological gender (e.g., *de vrouw* 'the woman' or *de koe* 'the cow') typically bear the corresponding lexical gender.

[1]We would like to thank the LiME-team of the Meertens institute for their comments on an earlier version of this paper, in particular we would like to thank Hans Broekhuis, Stef Grondelaers, Frans Hinskens, Khalid Mourigh, Piet van Reenen, Jos Swanenberg, and Ton van der Wouden,

Knowing the lexical gender of a word is part of an individual speaker's grammatical competence, whether explicit or not, although it can also be retrieved from prescriptive reference works, such as dictionaries or grammars. In dictionaries, all nouns are labelled as masculine (m), feminine (f), or neuter (n). Occasionally, a noun can be labelled as both masculine and feminine (m/f), and it is then called a 'hermaphrodite' (De Paepe and De Vogelaer 2008:1; Cornips and De Vogelaer 2009:8).

In the domain of adnominal modification masculine and feminine gender have coalesced into one category of common gender. This is a development that dates back to Middle Dutch (±1500 C.E.), when the gender distinction between masculine and feminine became less visible in the language due to the attrition and loss of suffixes that marked this very distinction (De Vogelaer and De Sutter 2011, Geerts 1966, Kraaikamp 2012, 2017). However, the distinction has been preserved in the southern Dutch dialects, i.e., the dialects of North Brabant and Limburg (as well as in Belgian Dutch), which mark adnominal modifiers of singular masculine nouns with the suffix *-e(n)*, e.g., *de-n hond* 'the-MASC dog' or *(e)en-en hond* 'a-MASC dog' (de Bont 1962, Hoppenbrouwers 1983, 1990, Stroop 1989, Weijnen 1971). Therefore, in general, the north of the Dutch-speaking area is regarded as more innovative when it comes to this linguistic change, while the south is regarded as more conservative (De Vos and De Vogelaer 2011) 9see also (De Troij et al. 2023) for more general discussion of this difference). For southern Dutch dialect speakers, the presence of the suffix acts as a linguistic cue to determine the lexical gender of the noun. However, even though this gender feature is one of the main characteristics of the southern Dutch dialects, recent research in the province of North Brabant has shown that the three-gender system is also subject to variation and change (Doreleijers et al. 2020, 2021).

Regardless of what the adnominal gender system looks like, however, all speakers of Dutch have to make a choice between the three genders (masculine, feminine, neuter) when selecting a pronoun, as in (1). The loss of visible gender distinctions affects the learnability of the system (Cornips and Hulk 2008:289, Cornips and De Vogelaer 2009:8, De Vos and De Vogelaer 2011, Mills 1986:112–115). Due to the lack of a clear distinction between masculine and feminine and a decreased knowledge of lexical gender, speakers regularly violate grammatical agreement, for example by referring to a neuter noun, such as *het boek* 'the book', with a masculine pronoun *hij*, *hem*, or *'m* 'he/him' Audring (2006:95). In addition, they sometimes resort to semantic default strategies, such as the use of *het* 'it' whenever the antecedent is not overtly marked as masculine or feminine (De Vogelaer and De Sutter 2011:203).

Previous research on the pronominal reference system in Dutch points to a change in the selection of pronouns in spoken language (see, among others, Audring 2006, 2009, Audring and Booij 2009, Cornips and De Vogelaer 2009, De Paepe and De Vogelaer 2008, De Vogelaer 2009, De Vos and De Vogelaer 2011, De Vos 2009, De Vos et al. 2021, Kraaikamp 2012, 2017, Romijn 1996). These studies have identified a shift from a lexical gender system, i.e., a system based on lexical gender like in German, to a semantic gender system like in English. In a lexical gender system, the choice of the pronoun is determined by the lexical gender of the antecedent noun, whereas in a semantic gender system the form of the pronoun is determined by the semantics of the antecedent noun (De Vos et al. 2021:31).

In this paper we provide new evidence of a pronominal gender shift in Netherlandic Dutch based on an experimental study with a large sample of speakers (n=10.119) from different age groups. Age effects are an important aspect of language change, because "in general, change in progress is typically implemented by successive generations of speakers pushing a phenomenon beyond levels observed in previous generations" (De Vos et al. 2021:39). The paper is organized as follows. Section 2 introduces the expected shift in pronominal gender marking and discusses the state of art on the topic. Section 3 details our methodology, while sections 4 and 5 present and discuss the main results.

2 Changing pronominal gender

Speakers of Dutch who must choose a pronoun and are unsure about the lexical gender of the noun the pronoun refers to, have roughly two strategies at their disposal (Audring 2021). The first is an avoidance strategy, which involves repeating the noun, replacing it by a synonym, or by a demonstrative pronoun.[2] Although this strategy is common, it is not the focus of the present paper. The other strategy involves speakers violating the grammatical requirement of agreement and instead choosing a pronoun based on the meaning of the antecedent noun (Audring 2006:91). For example, Audring (2006, 2009) has shown that in spoken Dutch the masculine pronoun *hij* 'he' and its variants '*ie*, *hem*, and '*m*, are used more often than would be expected based on the lexical gender of the antecedent noun. Specifically, the pronoun is not only used to refer to masculine persons, animals, and objects, but also to feminine animals and objects. This tendency is called masculinization (e.g., Geeraerts 1992, De Paepe and De Vogelaer 2008:4). In contrast, the feminine pronoun *zij* 'she' and its variants are only used for feminine persons. This too can involve a move away from a lexical gender system, for example when a neuter noun like *meisje* 'girl' is promominalized by *zij* or *ze* 'she' instead of *het* 'it' (De Vos et al. 2021:31). The neuter pronoun *het* 'it' on the one hand loses part of its function because it is superseded by masculine *hij* when referring to count nouns, but on the other hand it also gains additional functionality because it often replaces masculine and feminine pronouns when referring to substances, abstract entities, and events, i.e., when the noun is less concrete and less clearly bounded (e.g., Romijn 1996).

The shift from a gender system based on lexical agreement to one based on semantic agreement assumes an underlying cognitive hierarchy on which referents are placed according to the semantic properties that are attributed to them, i.e., their degree of individuation (Audring 2009:124). This Individuation Hierarchy, shown in (2), is a variant of the Animacy Hierarchy that emerged from typological research (Sasse 1993, Siemund 2008, Silverstein 1976).

(2) Humans > other animates (e.g., animals) > bounded objects > bounded abstract > specific mass > unspecific mass > unbounded abstract

One of the main assumptions of the Individuation Hierarchy is that there are no binary categories and no inherent properties of referents. Instead, speakers construe a referent as being unique or bounded (Audring 2009, Bouma 2018, De Vos et al. 2021:35). Previous research has shown that referents on the left-hand side of the hierarchy (e.g., entities with a biological gender) are more compatible with masculine or feminine pronouns, while referents more to the right on the hierarchy are more compatible with neuter pronouns. Or to put it differently: nouns that are highly individuated are referred to with masculine and feminine pronouns, whereas nouns that are less clearly individuated are referred to by the neuter pronoun (Audring 2009, De Vos et al. 2021:30, Hinskens et al. 2021). This process of changing pronominal gender is called resemanticization (Audring 2009, De Vos 2009, Wurzel 1986). In a semantic gender system, pronoun assignment is independent of the grammatical gender of the noun. However, oppositions in individuation, e.g., [concrete/abstract] or [count/mass], do play a role in pronoun assignment. A questionnaire study by De Vos and De Vogelaer (2011) finds that concrete count nouns show ca. 10% more lexical agreement than abstract count nouns, concrete mass nouns, or abstract mass nouns. These findings suggest a distinction between simple entities, i.e., tangible objects that have spatial integrity (Josefsson 2006:1352), and complex entities that lack such spatial integrity, including heterogeneous collections, semantic networks, and abstract notions (Romijn 1996:40–42). The results also show that a semantic and lexical system are not necessarily mutually exclusive but that they can co-occur. De Vos et al. (2021:59) formulate this as follows: "The alternation between lexical and semantic agreement depends on factors that facilitate or inhibit gender retrieval. [..] Simple entities have a high degree of so-called concreteness or imagery, making their lexical entry more accessible and the gender information more easily retrievable."

[2] Unlike personal pronouns, Dutch demonstratives only show a two-gender distinction between common (*deze*/*die*) and neuter (*dit*/*dat*) gender.

De Vos et al. (2021) weigh the influence of both structural and social factors on the choice of semantic and lexical agreement, based on the Belgian Dutch part of the Spoken Dutch Corpus. Their aim is to supplement Audring's (2009) account of the resemanticization of Dutch pronominal gender, by focusing on a region where the traditional three-gender system is still found. Their multivariate analysis reveals that most of the significant predictors are structural, for example the semantics of the referent (e.g., the use of neuter *het* 'it' is more likely for nouns with complex semantics), syntactic function, or the syntactic interval between antecedent and pronoun (see De Vos et al. (2021:48–56) for a detailed description of all factors). For the present study, however, social factors will be particularly of interest. The most important social factor so far uncovered in research on this topic is speech register (De Vos et al. 2021:55–56): semantic gender agreement is more often found in spontaneous speech settings, that is, when speakers pay less attention to their speech. This finding also emerges in a Twitter-based study from Bouma (2018) on semantically agreeing relative pronouns. Interestingly, (De Vos et al. 2021:30) argue that the effects they find support a psycholinguistic account "in which resemanticization is seen as a change from below (the level of social consciousness), caused by a hampered lexical access to noun gender". This suggests that semantically motivated pronouns qualify as indicators rather than markers (Labov 1972), implying that speakers are not aware of using them (De Vos et al. 2021:41).

Previous studies have also explored possible effects of gender and age. In line with Labov's (1990) finding that changes tend to be further advanced in women, female speakers use more semantic agreement than male speakers (Bouma 2018:152, De Vos et al. 2021:55). Furthermore, Audring (2009:169) reports that the semantic system is "asserting itself more strongly in the speech of younger generations". In her investigation of the Dutch part of the Spoken Dutch Corpus, speakers below 20 years of age use semantic agreement in over 70% of their utterances, whereas speakers aged 60 and above use it in only about 35% of their utterances. In his Twitter-based study, Bouma (2018:151) finds that the use of non-agreeing (i.e., semantic) relative pronouns is quite stable for users born before 1980, but steadily increasing for users born between 1980 and 2000, although this difference may be influenced by genre-specific differences, i.e., younger people being more accustomed to digital writing than older ones. For the Belgian Dutch language area, some questionnaire studies (De Vogelaer and De Sutter 2011, De Vos and De Vogelaer 2011) reveal significant differences between age groups, but the effect sizes are small, so practical significance may be limited.

The current study aims to get a clearer picture of gender and age effects by surveying a large sample of speakers. In addition, we applied a manipulation that increases salience, that is, speakers paying maximum attention to their pronoun choice, since a high level of salience has been found to yield a higher proportion of lexical (grammatical) agreement, i.e., conservatism (De Vos et al. 2021:59). The following research questions then arise:

(i) To what extent do full pronouns follow a pattern of lexical agreement in a formal and salient context?

(ii) What are these patterns of pronominal references?

(iii) How are these patterns related to the gender and age of participants?

The next section lays out the methodology used in the current study.

3 Methodology

This section is divided into two subsections. First, subsection 3.1 discusses the data collection method and the sentence completion task we used. Then, subsection 3.2 explains the way we analyzed the experimental data using Multiple Correspondence Analysis and Hierarchical Clustering.

3.1 Data collection

In order to investigate whether semantic agreement patterns also emerge when speakers are explicitly asked to select a pronoun within a given context, a sentence completion task was designed in collaboration with the Dutch popular scientific magazine *Quest*. Besides a paper magazine, Quest also operates a website.[3,4] On this website, people can read popular-scientific articles for free about all kinds of topics, including language, and also take quizzes on those topics. The idea of these quizzes (literally 'tests') is that people can test their knowledge of a particular subject in a short amount of time (maximally 10 minutes) and in an accessible way. Each test on the website ends with a score, challenging people to do their best, and also allowing them to compare their scores with others. Within this context, we developed a test on lexical gender in pronominal reference.[5] The test is accompanied by a short popular scientific article to explain the phenomenon of pronominal gender, but the participants were only referred to this article after completing the test, so as to not influence their answers.[6] Each participant could only take the test once. The test is still available online today, but the data collection for the current study took place from October 2019 to April 2020. The test was launched during the DRONGO language festival held on 25 and 26 October 2019 at Radboud University in Nijmegen. Visitors of the festival could join a 'live science lab' of the Meertens Institute, a research institute of the Royal Netherlands Academy of Arts and Sciences that studies the linguistic and cultural diversity in the Netherlands. In this lab, visitors could complete the test on the spot and interact with researchers of the institute. After the festival, the test remained available for several months and was advertized via social media platforms of Quest, the Meertens Institute, and the individual researchers, and through the digital magazine for Dutch linguistics and literature *Neerlandistiek.nl*.

The test consists of twenty sentence pairs in which a pronoun is omitted in the second sentence of every pair (see the Appendix for the full list of test items). Participants were presented with a forced choice design and had to choose one of three referential pronouns, i.e., masculine (option A), feminine (option B), or neuter (option C), such that the two sentences formed a logical and meaningful whole. The test did not include clitic or reduced pronouns next to full pronouns. This might have led to different results, especially since the reduced feminine pronoun *ze* 'she' could be regarded as more common to refer to objects compared to the full pronoun *zij*, see also Cardinaletti and Starke (1999). One of the sentence pairs is illustrated in (3). Participants could only choose one option and were required to make a choice for each item before they could proceed to the next question.[7]

(3) Als het regent draag ik een hoed. ____ hangt nu aan de kapstok.
if it rains wear I a hat hangs nu at the coat.rack
'When it rains, I wear a hat. ____ now hangs on the coat rack.'
A: *Hij* 'he'
B: *Zij* 'she'
C: *Het* 'it'

The test contains eight masculine nouns, six feminine nouns, and six neuter nouns. The nouns differ in their degree of individuation (concrete/abstract, count/mass, and animate/inanimate). An overview of all nouns that are included in the test is given in Table 1.[8]

In order to reach the higher level of salience, we made speakers aware of the pronoun selection process by

[3] https://www.quest.nl/

[4] Note that the privacy regulations of this forum do not allow us to collect geographical data of the participants. This is unfortunate since, as we explained in section 2 there is geographical variation in gender systems in the Dutch language area.

[5] The test can be found at the following url: https://tests.quest.nl/taal/de-het-meisje-weet-jij-woorden-mannelijk-vrouwelijk-onzijdig-zijn.

[6] The article can be found at the following url: https://www.quest.nl/maatschappij/taal/a29530668/de-het-woordgeslacht/.

[7] There is also a second part of the test, which focuses on relative pronouns, and whereby informants were shown sentence pairs and had to indicate which one(s) they found acceptable. In this paper we only analyze the first half of the test—the one focusing on pronominal gender—and so we leave the relative pronoun data out of the discussion.

[8] The variables, i.e. "gender", "(in)animacy", "count/mass", "concrete/abstract", have not been systematically combined, since we were restricted in the amount of test items we could include in the task.

NOUN	LEXICAL GENDER	(IN)ANIMATE	COUNT/MASS	CONCRETE/ABSTRACT
Hitte ('heat')	Feminine	Inanimate	Abstract	Mass
Informatie ('information')	Feminine	Inanimate	Abstract	Mass
Bibliotheek ('library')	Feminine	Inanimate	Concrete	Countable
Koe ('cow')	Feminine	Animate	Concrete	Countable
Oma ('grandmother')	Feminine	Animate	Concrete	Countable
Tante ('aunt')	Feminine	Animate	Concrete	Countable
Onzin ('nonsense')	Masculine	Inanimate	Abstract	Mass
Honing ('honey')	Masculine	Inanimate	Concrete	Mass
Auto ('car')	Masculine	Inanimate	Concrete	Countable
Hoed ('hat')	Masculine	Inanimate	Concrete	Countable
Schommel ('swing')	Masculine	Inanimate	Concrete	Countable
Tuin ('garden')	Masculine	Inanimate	Concrete	Countable
Boer ('farmer')	Masculine	Animate	Concrete	Countable
Hond ('dog')	Masculine	Animate	Concrete	Countable
Bed ('bed')	Neuter	Inanimate	Concrete	Countable
Boek ('book')	Neuter	Inanimate	Concrete	Countable
Hemd ('shirt')	Neuter	Inanimate	Concrete	Countable
Koekje ('cookie')	Neuter	Inanimate	Concrete	Countable
Vliegtuig ('airplane')	Neuter	Inanimate	Concrete	Countable
Paard ('horse')	Neuter	Animate	Concrete	Countable

Table 1: Overview of the nouns used in the sentence completion task

explicitly asking them to choose the grammatically correct pronoun and attaching a score to it in the form of an assessment task. Moreover, participants could only choose between full pronouns, not their reduced variants. Audring (2009:154) hypothesizes that speakers tend to overuse full pronouns in formal registers, and that they therefore tend to be more conservative, i.e., to show more lexical agreement. However, her data did not support this intuition, as full pronouns actually turned out to have a strong tendency towards semantic agreement (Audring 2009:216).

In addition to the linguistic questions, the test includes two background questions on the participant's age (open-ended question) and gender (multiple-choice question: male, female, other/not specified). These questions were compulsory for participants who wanted to see their score. In total, participants took about 10 minutes to complete the test. Participants who also wanted to receive their score by e-mail (or sign up for the Quest newsletter) could optionally leave their e-mail address. Quest did not share e-mail addresses with the researchers, so these are not stored for research purposes. All data were anonymous, i.e., not traceable to a specific person through personal information or IP addresses. [9]. In total, 10,353 participants filled out the questionnaire completely. In the next subsection we lay out the statistical analysis of our data.

3.2 Data analysis

We have analyzed the experimental data using a combination of Multiple Correspondence Analysis and Hierarchical Clustering, focusing first on the twenty words in Table 1 and how they group together, and then on our pool of participants, to see what strategies they use to assign a pronominal gender to these twenty words.[10] The starting point for the first part of the analysis is a raw data table a small portion of which is shown in Table 2.

	2957369	2574003	2956679	...	GENDER	CONC-ABS	COUNT-MASS	(IN)ANIM
hoed	Het	Hij	Zij	...	M	CONC	COUNT	INANIM
tante	Zij	Zij	Zij	...	F	CONC	COUNT	ANIM
auto	Hij	Hij	Hij	...	M	CONC	COUNT	INANIM
tuin	Het	Hij	Hij	...	M	CONC	COUNT	INANIM
hemd	Het	Het	Zij	...	N	CONC	COUNT	INANIM
...	...	...	...	...	...	...	...	...

Table 2: The raw data table that served as input for the analysis (partly)

The data partially represented in Table 2 form a 20×10,357 matrix, with each row representing one of the twenty words listed in Table 1. The first 10,353 columns represent our participants. Each participant was randomly assigned a seven-digit identification number, and the cell values are one of the three multiple choice options offered in the questionnaire (see the example in (3)): *Hij* 'he', *Zij* 'she', or *Het* 'it'. The final four columns represent the information that was also listed in Table 1: the lexical gender of the nouns, whether they are concrete or abstract, count or mass, and animate or inanimate. On this table we performed a Multiple Correspondence Analysis (MCA), with the final four columns acting as supplementary variables (for general discussion, see Greenacre 2007 and Levshina 2015:chapter 19). The output of MCA is a low(er)-dimensional representation of

[9] As the data for this study were collected through Quest's online platform, the data collection was according to their privacy statement. This statement can be found via the following url's: https://www.quest.nl/tech/wetenschap/a2605908o/over-quest-test-nederland/ and https://www.hearst.nl/privacyverklaring/ . The researchers did not have access to the raw data, only to the processed data. The data were collected and analyzed within the NWO (Dutch Research Council) funded project 'Changing gender: language variation and change in gender marking in Dutch dialects', PGW.19.018

[10] All calculations were carried out in R (R Core Team 2014) using the FactoMineR-package (Husson et al. 2014).

the data set,[11] which makes it possible to visualize and subsequently interpret how the nouns pattern in terms of pronominal reference and what role the supplementary variables play in determining this patterning. The final step in our analysis involves Hierarchical Clustering (HC), which divides the set of twenty nouns into a number of subsets such that these subsets reflect common behavior in the questionnaire. We perform this clustering on the outcome of the MCA (Husson et al. 2010, 2011). In other words, we use the coordinates of the twenty nouns in the (reduced) MCA-space as the basis for clustering them into groups. All results of both the MCA and the HC are discussed in section 4.

The analyses just described can provide insight into which (types of) nouns cluster together in our experiment, but they don't yet provide any information about which pronouns our participants actually use to refer to these nouns. In order to answer that question, we performed a second MCA and HC analysis on our data set, but this time focused on the participants, and using their age, gender, and test score as supplementary variables. This means our initial raw data set looks slightly different. A small portion of it is shown in Table 3.

	hoed	informatie	boek	...	AGE	GENDER	SCORE
2572648	Hij	Zij	Het	...	24	female	17
2573832	Hij	Het	Hij	...	26	female	13
2957372	Het	Het	Hij	...	17	female	11
2957380	Hij	Zij	Het	...	16	female	17
...	...	...	...	...	...	...	...

Table 3: The raw data table that served as input for the second half of the analysis (partly)

The data table underlying Table 3 measures 10,119 rows by 23 columns. Each participant is represented by a separate row.[12] The first twenty columns are the twenty words of the questionnaire, with cell values identical to the ones in Table 2, and the final three columns contain the age of the participant, their gender, and their test score. The age of the participants ranges from 11 to 99, with a mean of 44.56 (median: 44.00) and a standard deviation of 17.58. In order to facilitate the visualization of the data set and the analysis of the results, we split up the participants into seven different age groups. The range and size of these groups is given in Table 4.

age range	**number of participants**
10 to 19	654
20 to 29	1980
30 to 39	1708
40 to 49	1541
50 to 59	1788
60 to 69	1588
70 to 99	860

Table 4: Age ranges of our participants

As far as their gender is concerned, 6004 of the participants are female (59%) and 4115 male (41%). Finally, the SCORE-variable in Table 3 is a number that theoretically can range from 0 to 20 and that indicates the degree

[11] Given that the first three dimensions combined account for 68.4% of the variance in the original data set, we will focus on those three dimensions in the discussion of the results in section 4.

[12] For this part of the analysis, we excluded participants with a (self-indicated) age lower than 10 or higher than 100. In addition, we excluded the 139 participants who chose the option 'other/not specified' in the question regarding their gender.

to which the pronoun choices made by the participant reflect the Standard Dutch lexical gender of the relevant nouns, i.e. it is the sum of the prescriptively correct answers over the twenty stimuli. As such these scores provide us with a measure of the conservatism of the participants' grammars. In our data set, the lowest score is 4 and the highest 20. The mean is 13.59 (median: 14.00) and the standard deviation 3.07. As shown in Table 5, apart from the very high and very low scores, the participants are fairly evenly distributed over the different scores.

score	number of participants	score	number of participants
4	1	13	1099
5	5	14	1108
6	20	15	1133
7	99	16	917
8	292	17	796
9	569	18	620
10	820	19	402
11	1017	20	174
12	1047		

Table 5: Scores obtained by the participants on the test

In the next section we discuss the results of the analyses described above.

4 Results

4.1 Noun clusters

We first present the results of the MCA performed on the data table partially represented in Table 2, i.e., the part of the analysis that focuses on discovering commonalities among the twenty nouns of the experiment. The first two MCA-dimensions are visualized in the plot in Figure 1.

This plot is a two-dimensional representation of the twenty nouns from Table 1. The closer two nouns are to one another in the plot, the more our participants referred to them using the same pronoun (regardless of whether that pronoun is *hij* 'he', *zij* 'she', or *het* 'it'). Nouns that are far apart are typically not referred to by the same pronoun. The question now is if we can detect natural classes in the groups of nouns that cluster together, and this is where the supplementary variables come in. Recall from (the discussion of) Table 2 that we also encoded the lexical gender of the noun, as well as a number of additional properties. That information was not used to construct the plot in Figure 1, but it can be superimposed on it, in various ways. First we can calculate, for each supplementary variable and each MCA-dimension, the squared correlation ration (η^2). This is a value between 0 and 1 that represents the proportion of variance explained by that supplementary variable on that MCA-dimension. The η^2-values for the four supplementary variables and the first three MCA-dimensions are given in Table 6.

As is clear from the table, lexical gender plays an important role in accounting for the variation, especially on the first two dimensions, while the count/mass distinction is the most important driver of variation on the third dimension.[13] These findings can be further corroborated when we color-code the plot in Figure 1 according to the values of the supplementary variables. Consider in this respect the plot in Figure 2.

[13]It is hard to find absolute measures for η^2 to determine the size of the effect. Some authors cite Cohen 1962, in which case an η^2-

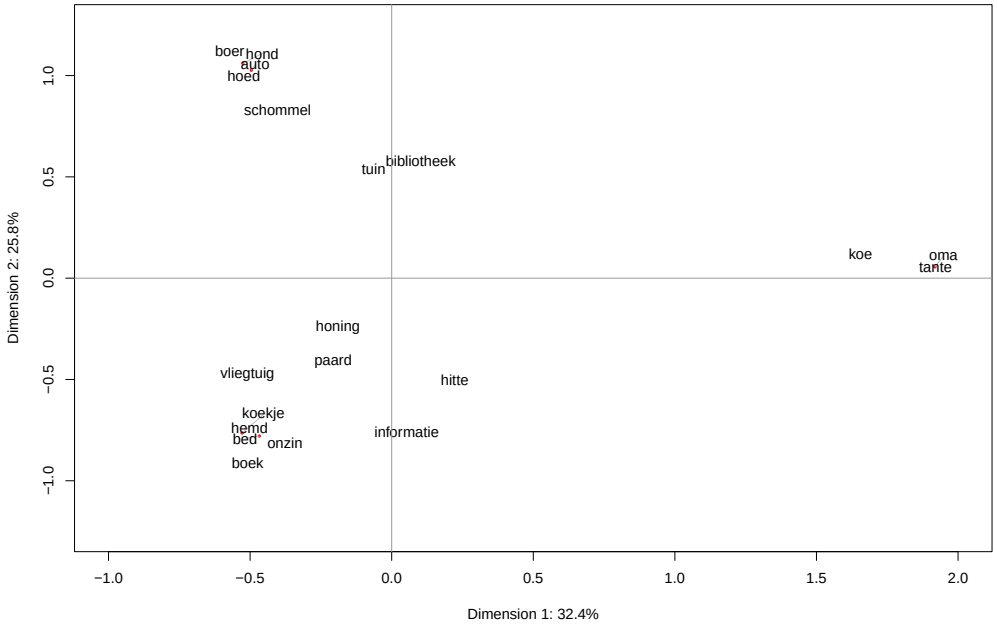

Figure 1: First two dimensions of the MCA

	Dim 1	Dim 2	Dim 3
GENDER	0.63	0.52	0.43
CONC/ABS	0.00	0.16	0.46
COUNT/MASS	0.00	0.16	0.69
(IN)ANIM	0.32	0.09	0.14

Table 6: Squared correlation ratio for the four supplementary variables and the three MCA-dimensions

This is the same plot as the one in Figure 1, but color-coded according to the lexical gender of the nouns: feminine nouns are marked in black, masculine nouns in red, and neuter nouns in green. Note how this color-coding closely matches the grouping of the nouns in the plot: black nouns are on the right-hand side of the plot [14], green ones occupy the lower left quadrant, and red ones the upper left quadrant—with two exceptions, which we return to below. This shows that in choosing a pronoun to refer to these nouns, our participants strongly relied on their lexical gender. Using the same line of reasoning, we can now also look at the third dimension of the MCA. Consider in this respect the plot in Figure 3.

This plot represents the second and third dimension of the MCA, and this time we have color-coded the nouns according to whether they are count or mass nouns, i.e. the supplementary variable with the highest η^2-value on this dimension (see Table 6 above). Again it is clear that the color-coding tracks the grouping in the plot: all mass nouns (marked in red) are grouped together in the lower left quadrant of the plot. This also explains the unexpected position of the nouns *honing* 'honey' and *onzin* 'nonsense' in Figure 1: while masculine, these nouns are also mass, and that is what is driving them downwards on the plot.

A third and final way to superimpose the information provided by the supplementary variables on the plots

value of 0.0099, 0.0588, and 0.1379 would correspond to a small, medium, and large effect respectively, but see Richardson 2011 for critical discussion. At any rate, it should be clear that the η^2-values of lexical gender and (on the third dimension) concrete/abstract and count/mass should count as large.

[14] We will get back to the fact that the feminine nouns on the right-hand side of the plot are all animate below

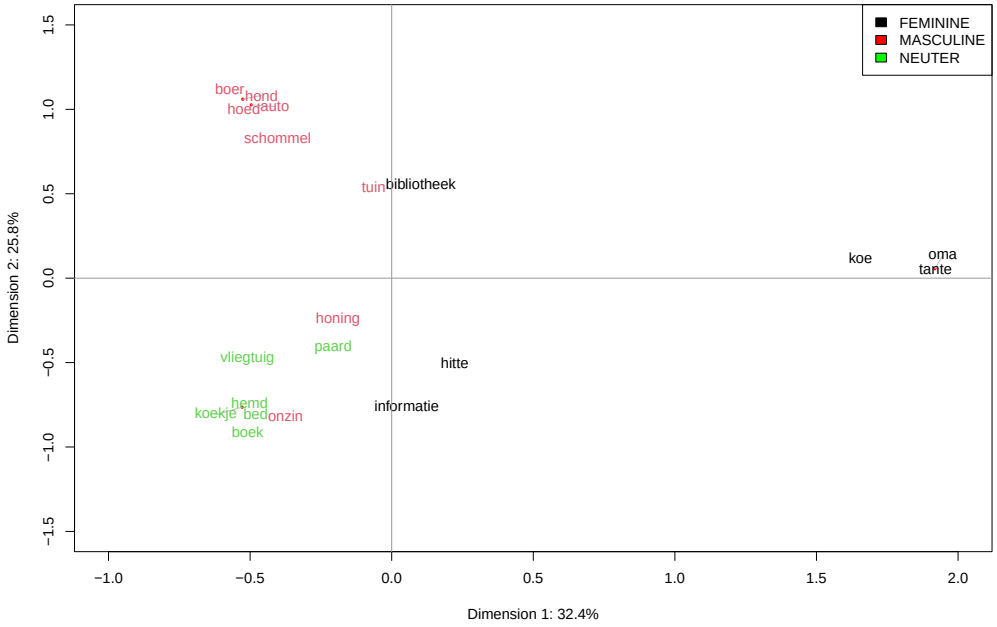

Figure 2: First two dimensions of the MCA color-coded according to lexical gender

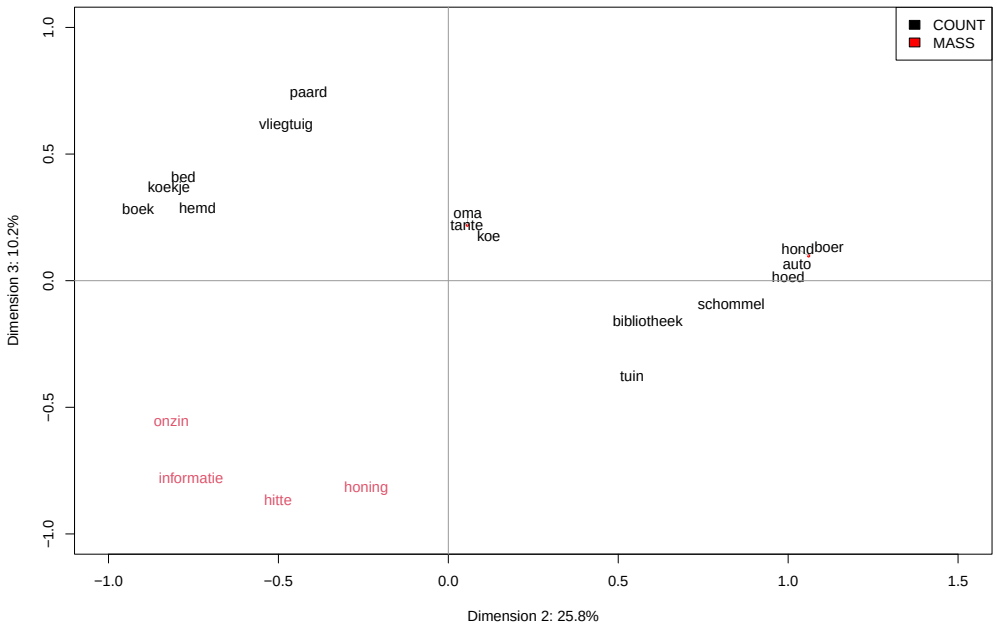

Figure 3: Second and third dimension of the MCA

generated by the MCA is by directly plotting the values of the supplementary variables. Just as each of the twenty words from Table 1 can be plotted on the MCA-dimensions, so can the values of the supplementary variables. The result is shown in the plots in Figure 4. This way of representing the data reveals a number of additional insights. First, it illustrates an effect from Table 6 that we have so far not commented on, namely the role of animacy. While all feminine nouns are to the right of the *y*-axis in Figure 2, there is still a clear division

between *koe* 'cow', *oma* 'grandmother', and *tante* 'aunt' all the way on the right, and *bibliotheek* 'library', *hitte* 'heat', and *informatie* 'information', which are situated more in the center of the plot. The left-hand plot in Figure 4 shows that this is the effect of animacy: it is not so much feminine nouns in general that cluster together on the first dimension, but *animate* feminine nouns. A second thing to note in Figure 4 is the fact that the values CONCRETE and COUNT occupy a position close to the origin in both plots. This shows that these values are not very distinctive in our data set. In other words, our participants rarely used the same pronoun for all concrete nouns, or all count nouns (as opposed to, say, the mass nouns or the animate feminine nouns).

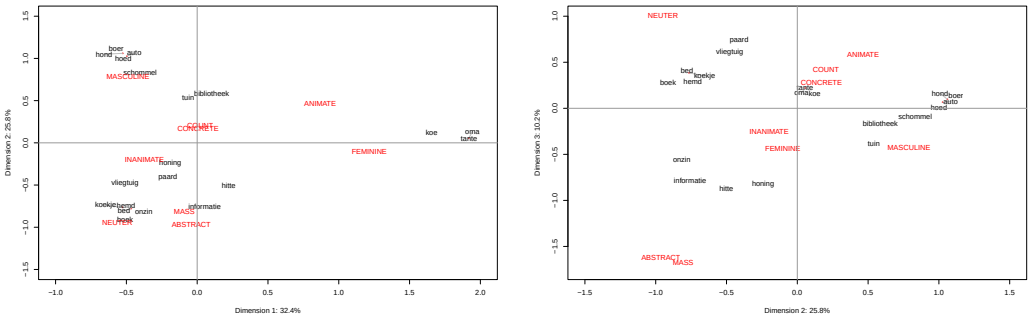

Figure 4: MCA with the values of the supplementary variables also plotted. The first two dimensions are shown on the left, the second and third on the right.

The trends we have just observed are further corroborated if we look at the outcome of the hierarchical clustering analysis. Applying Hierarchical Clustering to the outcome of MCA yields the following four clusters:[15]

(4)

cluster 1	**cluster 2**	**cluster 3**	**cluster #4**
hoed 'hat'	*tante* 'aunt'	*hemd* 'shirt'	*informatie* 'information'
auto 'car'	*oma* 'grandmother'	*bed* 'bed'	*onzin* 'nonsense'
tuin 'garden'	*koe* 'cow'	*boek* 'book'	*hitte* 'heat'
bibliotheek 'library'		*koekje* 'cookie'	*honing* 'honey'
boer 'farmer'		*vliegtuig* 'airplane'	
schommel 'swing'		*paard* 'horse'	
hond 'dog'			

This clustering further confirms our earlier analysis of the data, in particular the role played by gender, the count/mass-distinction, and animacy. Cluster 2 contains all and only animate feminine nouns, cluster 3 all and only neuter nouns, cluster #4 all and only non-neuter mass nouns, and cluster 1 contains all remaining nouns, i.e., the masculine nouns (whether animate or not) and the inanimate feminine nouns.

4.2 Participant clusters

As already hinted at in the previous section, the results we have just discussed paint a clear picture about which (types of) nouns cluster together in our experiment, but they don't yet provide any information about which pronouns our participants use to refer to these nouns. For instance, the MCA and subsequent cluster analysis have revealed that the four nouns in cluster #4 in (4) are typically referred to by the same pronoun, but which pronoun that is, and whether there is variation among our participants in the choice of pronoun is still an open question. This is why we now turn to the second MCA and cluster analysis described in the previous section, i.e.,

[15] The average silhouette width of these clusters is 0.62, and none of the clusters has a silhouette width smaller than 0.5.

the one that focuses on the participants, and uses their age, gender, and test score as supplementary variables. As a first visualization of the raw data, consider the bar plots in Figure 5. They show, for each of the twenty words, which percentage of our participants chose which pronoun. These plots are already quite informative. For instance, they show that for nouns like *tante* 'aunt' and *oma* 'grandma' there is much greater agreement between the participants than for nouns like *vliegtuig* 'plane' or *tuin* 'garden'.

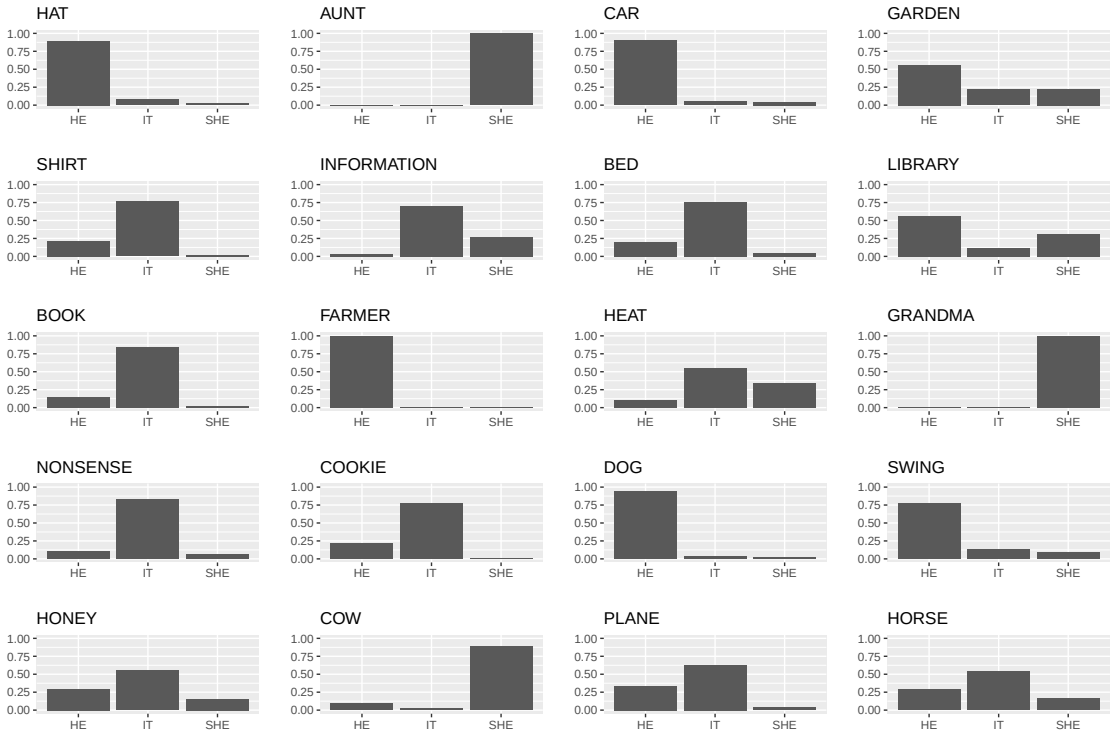

Figure 5: Bar plot representation of the raw data

In order to make these intuitions about the degree of inter-speaker agreement more quantifiable, we adopt and adapt the notion of 'internal onomasiological uniformity' (IOU) from Geeraerts et al. (1999:46). They define IOU as follows:

(5) $I_Z(Y) = \sum_{i=1}^{n} F_{Z,Y}(x_i)^2$

In this formula $I_Z(Y)$ represents the internal onomasiological uniformity for concept Z in data collection Y, x_1 to x_n are possible expressions of concept Z in data collection Y, and $F_{Z,Y}(x)$ is the relative frequency of x in Y to express Z. Geeraerts et al. (1999) use this concept to measure the degree to which there is uniformity in the terms speakers use to refer to a certain concept: when there is a clearly dominant term the value of $I_Z(Y)$ will be high, whereas when several terms are more or less equally popular, $I_Z(Y)$ will be much smaller. We can now use this same concept to quantify the degree of uniformity in the bar plots in Figure 5: when most speakers converge on the same pronoun (as in the case of *tante* 'aunt' or *oma* 'grandma'), the value will be high, whereas in cases like *vliegtuig* 'plane' or *tuin* 'garden' it will be much smaller).[16] The actual IOU-values are given in Table 7.

A comparison between these scores and the classes defined by the Individuation Hierarchy in (2)—repeated below as (6)—reveals clear similarities. First, the nouns with the highest uniformity scores are those that are highly individuated. Nouns like *tante* 'aunt', *oma* 'grandma' and *boer* 'farmer' all refer to humans (i.e., biological gender), which occupy the extreme left edge of the hierarchy. In almost all cases, these nouns were referred to with feminine ('aunt', 'grandma') and masculine ('farmer') pronouns.

[16] Many thanks to Stef Grondelaers (p.c.) for suggesting this measure to us.

NOUN	UNIFORMITY SCORE	NOUN	UNIFORMITY SCORE
AUNT	99.37	SWING	63.35
GRANDMA	99.19	SHIRT	63.06
FARMER	98.66	BED	61.57
DOG	88.31	INFORMATION	56.07
CAR	82.33	PLANE	51.04
HAT	80.77	HEAT	43.49
COW	79.02	LIBRARY	43.14
BOOK	72.77	HONEY	41.31
NONSENSE	70.95	GARDEN	41.03
COOKIE	64.78	HORSE	40.73

Table 7: Overview of the nouns used in the sentence completion task

(6) Humans > other animates (e.g., animals) > bounded objects > bounded abstract > specific mass > unspecific mass > unbounded abstract

Moving slightly to the right of the Individuation Hierarchy, the next highly individuated class includes other animates, e.g., animals. Indeed, the noun *hond* 'dog' is in fourth place in the ranking of uniformity scores in Table 7, since this noun is referred to by the masculine pronoun in almost all cases. The same applies to the noun *koe* 'cow', which is almost exclusively assigned a feminine pronoun. However, the opposite is true for the neuter noun *paard* 'horse', as this noun has the lowest uniformity score in the ranking. In almost half the cases, 'horse' is assigned a common pronoun, more often masculine than feminine. This finding goes against the lexical agreement system (i.e., a neuter noun should be referred to with a neuter pronoun), but it is not unexpected from the point of view of the Individuation Hierarchy, which assumes that highly individuated nouns are more compatible with masculine or feminine pronouns.

In the middle of the Hierarchy we find the class of bounded objects, such as *hoed* 'hat' or *boek* 'book'. This class of count nouns is less individuated than the categories of humans and other animates, but more than mass nouns. Here, the masculine nouns *auto* 'car' and *hoed* 'hat' have a relatively high uniformity score, as in the vast majority of cases these nouns are assigned a masculine pronoun. However, this does not apply to the masculine noun *schommel* 'swing'. Compared to 'car' and 'hat', the noun 'swing' is increasingly referred to with a feminine or neuter pronoun. Within the tendency of masculinization and the relatively high degree of individuation, this is unexpected. Still, the majority of the pronouns used are masculine.

Within the class of count nouns, the masculine noun *tuin* 'garden' and the feminine noun *bibliotheek* 'library' have the lowest uniformity scores. In about half the cases the noun 'garden' is referred to with a feminine or neuter pronoun. Particularly the neuter pronoun is unexpected here, as 'garden' is a count noun. However, at the same time, the noun indicates a place (in contrast to 'book' or 'cookie', for example) and could therefore be regarded as more abstract and less individuated, causing variation in pronoun assignment. The feminine noun 'library' also shows low uniformity. In about half the cases it is referred to with a masculine pronoun. We can thus conclude that nouns indicating a place seem to be more subject to variation than countable bounded objects (see also Audring 2009:69).

The uniformity scores of neuter count nouns like *boek* 'book', *koekje* 'cookie', *hemd* 'shirt', and *bed* 'bed' are clearly lower than those of their non-neuter counterparts. Most of the time a neuter pronoun is assigned, but the variation increases, particularly the use of masculine pronouns to refer to these nouns. This fits in with the idea that the neuter pronoun loses part of its function due to masculinization, i.e., the masculine pronoun taking

over the position of neuter pronouns when referring to count nouns. This shift seems even further advanced in the neuter noun *vliegtuig* 'plane', with more than a quarter of the pronouns assigned being masculine.

Moving further to the right on the Hierarchy, we arrive at the class of specific mass nouns. In our data set, the masculine noun *honing* 'honey' has one of the lowest uniformity scores. The high number of neuter pronouns fits within the process of resemanticization and the idea that referents on the right-hand side of the Hierarchy are lowly individuated and more compatible with the neuter pronoun regardless of their lexical gender.

At the extreme right edge of the Hierarchy we find the class of unspecific mass nouns and unbounded abstracts. The noun *onzin* 'nonsense' has a relatively high uniformity score. Looking at the pronoun assignment in Table 5, it is striking that in more than three quarters of the cases this common noun was referred to with the neuter pronoun, thus breaking with lexical agreement in favor of semantic agreement. Although the other lowly individuated nouns *informatie* 'information' and *hitte* 'heat' have lower uniformity scores, both nouns also clearly have the neuter pronoun as the most frequently chosen pronoun.

In summary, the uniformity scores in Table 7 track the Individuation Hierarchy in (6) quite well. They generally show the highest scores for animates, in which semantic gender also matches lexical gender. If the noun's semantics clashes with its grammatical gender, semantic agreement becomes more likely and uniformity scores decrease. This was the case with animals like *paard* 'horse', places like *tuin* 'garden', specific mass nouns like *honing* 'honey', and unbounded abstract nouns like *hitte* 'heat'. In the middle of the Hierarchy, semantic agreement occurs less frequently, but there are some indications of an ongoing process of masculinization with the masculine pronoun sometimes taking the position of neuter or feminine pronouns. Yet another indication of pronominal gender shift is the semantic agreement among nouns with neuter diminutive or feminine derivational suffixes from which the lexical gender could actually be read off the linguistic form directly, e.g. *-je* in *koekje* 'cookie' or *-te* in *hitte* 'heat'.

4.3 Combining participant clusters and noun clusters

What neither the plots in Figure 5 nor the numbers in Table 7 show, however, is the patterns of variation that underlie the data, i.e. are there groups of speakers that we can identify based on their pronominalization strategies? This is why we now turn to the results of the MCA and cluster analysis. The cluster analysis allows us to split up the 10,119 participants into three clusters, the sizes of which are shown in Table 8.[17] It is the properties of these three groups that we will examine more closely in what follows.

cluster number	number of participants
1	3525
2	5477
3	1117

Table 8: Size of the three participant clusters

We first examine the interaction between age and cluster membership. This relationship is visualized in Figure 6. The figure in this plot represents the three clusters on the *y*-axis in increasingly darker shades of gray. The *x*-axis represents the seven age ranges, and the width of the bars is proportional to the size of that age group (see Table 4). The height of the bars shows the percentage of participants from a particular cluster in that age range. What the figure shows is that cluster 1 skews towards the older age ranges, while in clusters 2 and 3 younger participants are overrepresented.

This is confirmed by the plot in Figure 7. This plot represents the first two dimensions of the MCA, color-coded according to the three clusters revealed by the cluster analysis. Superimposed on this we have plotted the

[17] The silhouette widths are 0.33, 0.21, and 0.34 respectively (average silhouette width: 0.26).

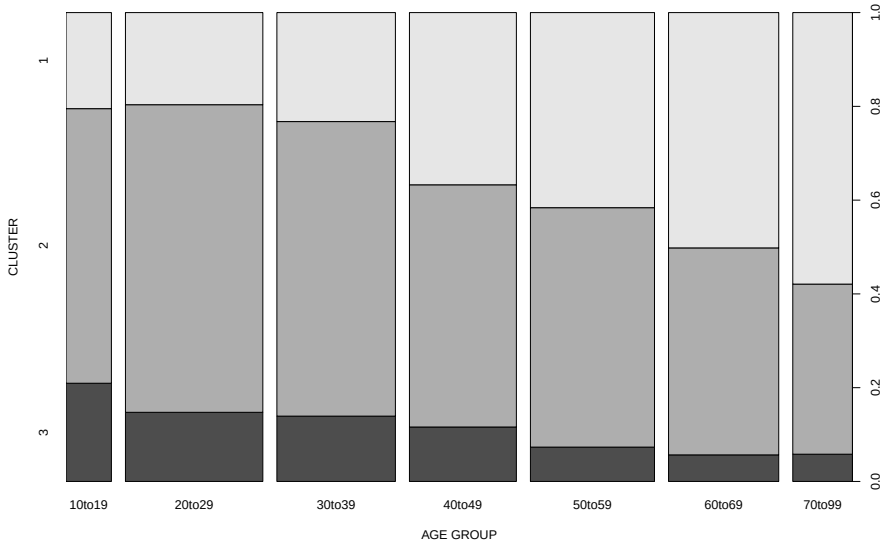

Figure 6: Interaction between age ranges and cluster membership

seven age ranges. As can clearly be seen from the plot, the age variable aligns quite well with the first dimension of the MCA, with higher age ranges being represented on the left-hand side of the plot, and lower age ranges on the right-hand side. This same horizontal dimension is also a key factor in distinguishing the three clusters, with cluster 1 exclusively to the left of the *y*-axis, cluster 3 to the right, and cluster 2 in an intermediate position.[18]

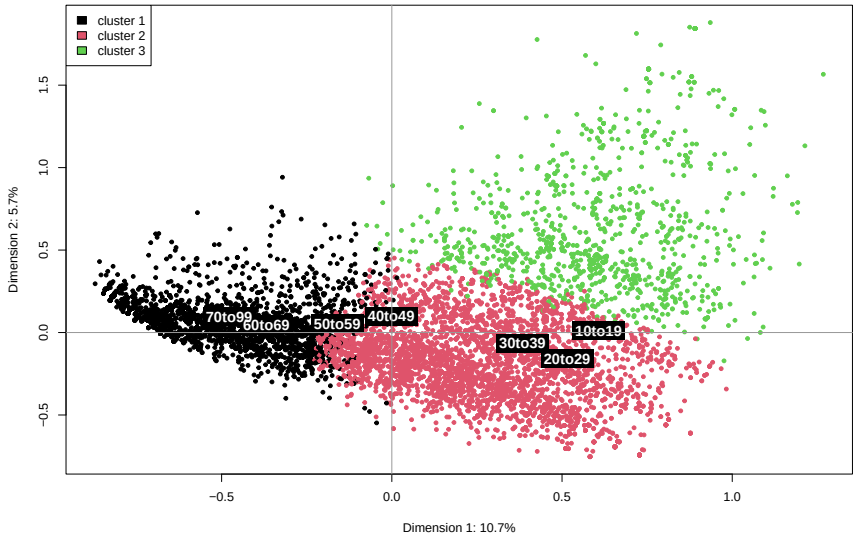

Figure 7: First two dimensions of the MCA color-coded according to cluster and with age ranges superimposed

Let us now focus on the interaction between gender and cluster membership. The plot in Figure 8 parallels

[18]The supplementary variable AGE has an η^2 of 0.139 on the first dimension (and 0.008 on the second), but this is not that informative, because the value of η^2 for the combination of a dimension and a particular categorical variable is sensitive to the number of values this variable can have: the higher the number of possible values, the higher the value of η^2 (see Richardson (2011) for discussion). In other words, variables with a high number of possible values tend to have inflated η^2-scores. In order to circumvent this, we also ran the MCA with age as a quantitative supplementary variables—i.e. without splitting it up into age ranges—and that variable does indeed come out as inversely correlated with the first MCA-dimension, with higher ages corresponding to lower values on the *x*-axis.

the one in 6 but it substitutes the seven age cohorts for two genders. As is clear from eyeballing the plot, gender does not seem to play a role in accounting for the variation in our data set. This is also confirmed by the MCA, in two ways. First, the η^2 of the supplementary variable GENDER is 0.013 on the first MCA-dimension and 0.008 on the second, showing that the effect is very small indeed. Secondly, if we plot the values for this supplementary variable on the first two dimensions, as is done in Figure 9, it is clear that both of them are very close to the origin of the plot and hence do not carry a lot of weight. At most there is a slight preponderance of female speakers in cluster 2—note how "female" is situated in the lower right quadrant in Figure 9, the quadrant in which cluster 2 is dominant—which was also suggested by the bar plots in Figure 8.

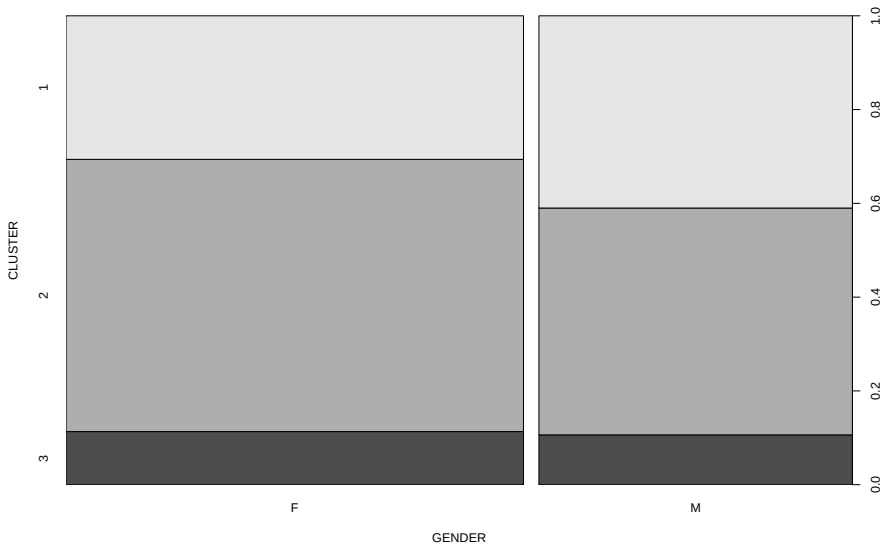

Figure 8: Interaction between gender and cluster membership

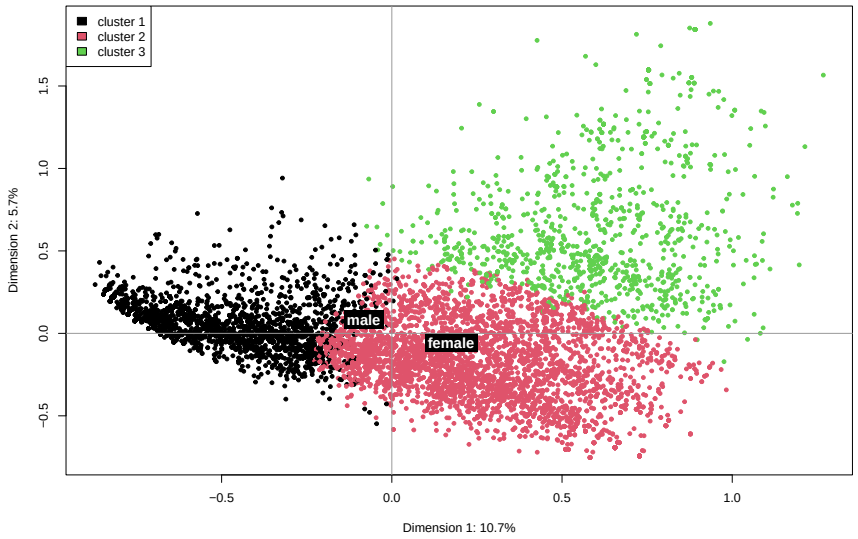

Figure 9: First two dimensions of the MCA color-coded according to cluster and with gender superimposed

Finally, we turn to the interaction between test score and cluster membership. The result of this analysis is represented in Figure 10. In this graph, the effect of the supplementary variable is much more pronounced than

in the previous two figures: high scores are almost exclusively found in cluster 1, with cluster 2 dominating the mid to low range, and cluster 3 containing almost exclusively low scores. The effect of test score can also be seen in the outcome of the MCA. In Figure 11 we plot the first two MCA-dimensions color-coded according to test score. The graph clearly shows that red/high scores dominate on the left (where cluster 1 is situated, see Figure 7), and low/green scores on the right.[19]

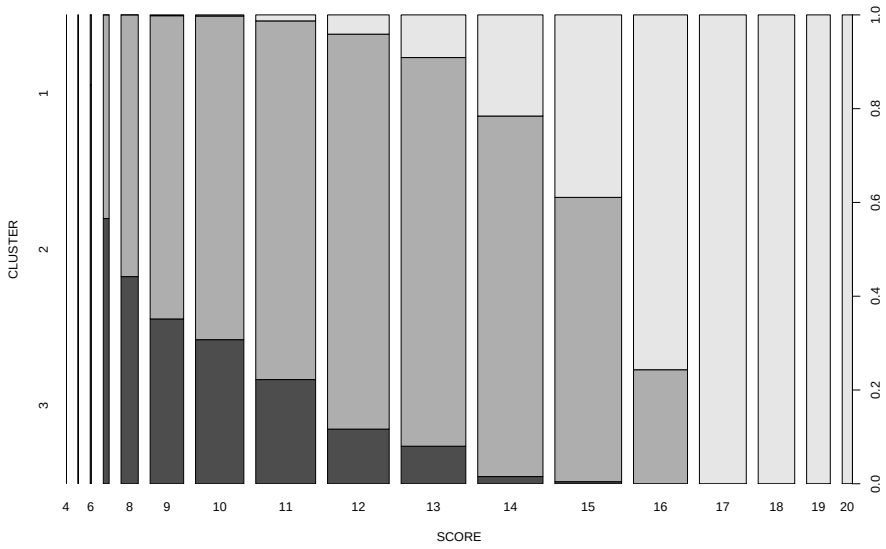

Figure 10: Interaction between test scores and cluster membership

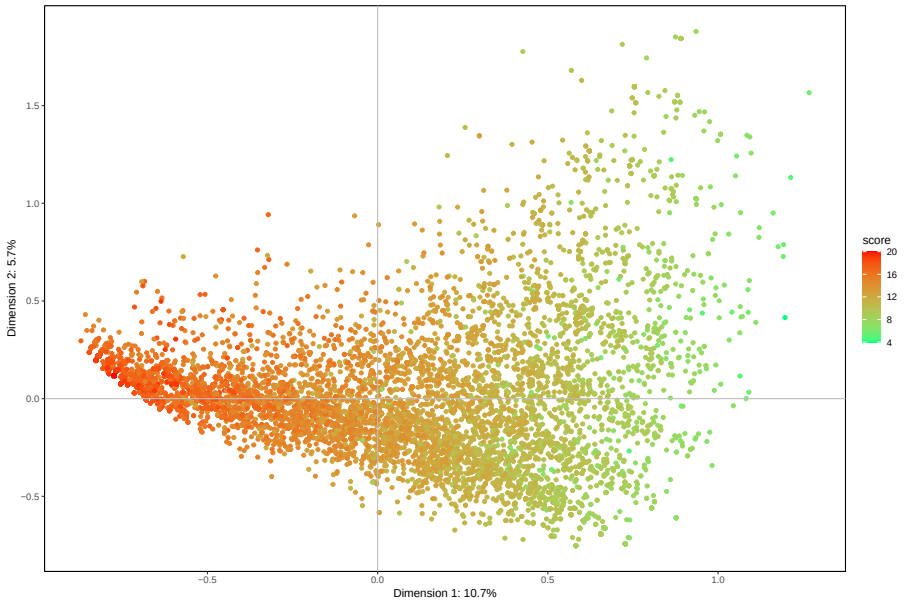

Figure 11: First two dimensions of the MCA color-coded according to test score

The results we have just discussed provide the following characterisation of our participant pool: there is an older, more conservative group of speakers whose judgments adhere closely to the Standard Dutch norm, and

[19]The η^2 of the variable SCORE is 0.883 on the first dimension and 0.030 on the second, but as pointed out in fn18, these values are artificially high because SCORE has 17 possible values. We have also run the MCA with SCORE as a quantitative supplementary variable, however, and it comes out as very strongly (inversely) correlated with the first MCA-dimension.

there is a younger, more innovative group of speakers. In between the two extremes we find a continuum, both in terms of age and in terms of degree of conservatism. In order to get a clearer view of the types of innovation that we find among our (younger) participants, we can look at the type of plot shown in Figure 5, but this time split up per cluster. This is shown in Figure 12.

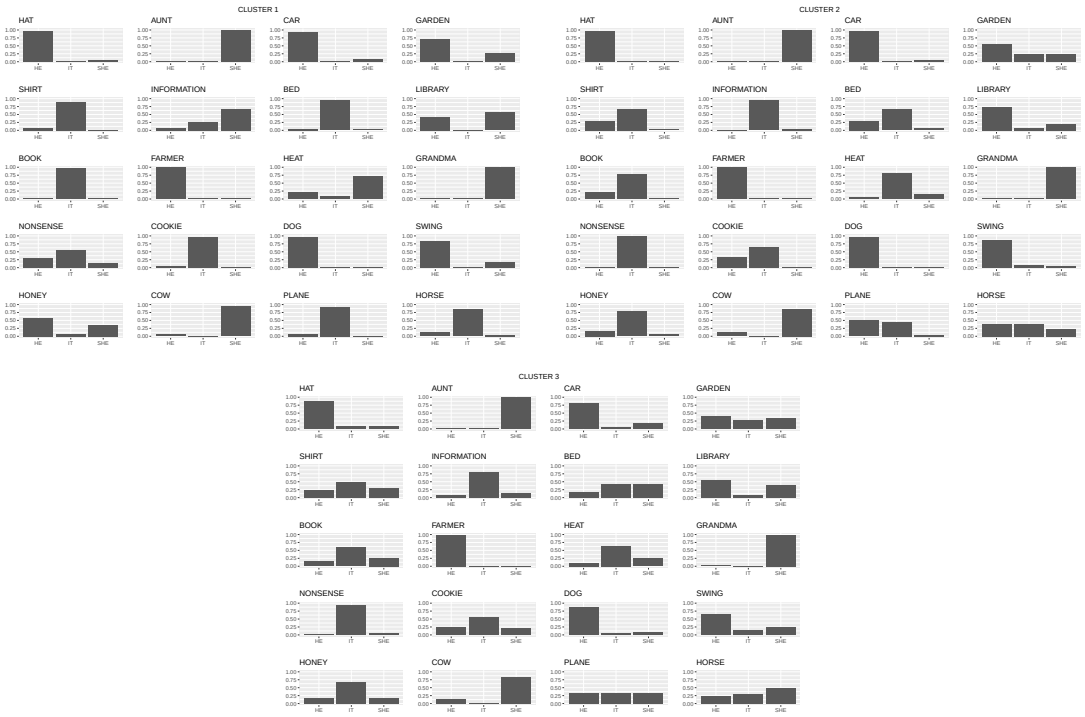

Figure 12: Bar plot representation of the data per cluster

Cluster 1 corresponds to the highest test scores and is thus closest to the lexical agreement system. With the exception of the noun *onzin* 'nonsense', which is mostly assigned a neuter pronoun, pronoun selection in this cluster takes place on the basis of the lexical gender of the noun. For example, the other abstract and mass nouns do not show any indications of semantic agreement. In fact, the nouns that are most variable within this cluster, such as *honing* 'honey' and *hitte* 'heat', reveal uncertainty about which of the two common genders should be assigned rather than switches between with common and neuter gender.

Clusters 2 and 3 clearly differ from cluster 1, in that abstract and mass nouns such as *informatie* 'information', *hitte* 'heat', and *honing* 'honey' are mostly assigned a neuter pronoun. This is a clear indication of a shift from a lexical to a semantic agreement system in the younger, more innovative speaker groups. In addition, compared to cluster 1, uniformity decreases for the locational nouns *garden* 'garden' (more references with a neuter pronoun) and *bibliotheek* 'library' (more references with a masculine pronoun). Also, there is a higher number of occurrences of a masculine pronoun when referring to neuter concrete count nouns, i.e., *hemd* 'shirt', *bed* 'bed', *boek* 'book', *koekje* 'cookie', and *vliegtuig* 'plane'. This points to an increasing degree of masculinization. The neuter animal noun *paard* 'horse' is also increasingly referred to with a non-neuter pronoun, which is consistent with its high degree of individuation according to the Individuation Hierarchy. In cluster 3, however, the number of occurrences of a feminine pronoun increases, especially in reference to neuter concrete countable objects such as 'book' and 'bed'. This is unexpected, as previous research suggested that the use of the feminine pronoun *zij* 'she' is limited to feminine persons in a semantic agreement system.

In sum, clusters 2 and 3 show more semantic agreement and instances of masculinization than cluster 1, thereby accounting for the lower test scores. However, cluster 3 can be distinguished from cluster 2 based on the (slightly) higher proportion of feminine pronouns in reference to non-feminine nouns.

5 Discussion

5.1 Summary

The results of the Multiple Correspondence Analyses and Hierarchical Clustering show that studying a large sample of speakers helps consolidate insights about pronominal gender shift in Dutch. Below we will come back to the research questions we started this article with.

(i) To what extent do full pronouns follow a pattern of lexical agreement in a formal and salient context?

(ii) What are these patterns of pronominal references?

The data set makes it possible to detect natural classes in the groups of nouns that cluster together, such as animate feminine nouns and non-neuter mass nouns. Participants not only strongly rely on the lexical gender of the nouns in choosing a referential pronoun (the lexical agreement system), but they also use semantic nominal properties such as animacy, countability, and concreteness in guiding their pronominalization strategies (the semantic agreement system). Whereas the values CONCRETE and COUNTABLE did not play a clearly identifiable role in pronoun assignment, i.e., participants rarely use the same pronoun for all concrete nouns or all count nouns, the opposite was true for the values ANIMATE, UNCOUNTABLE/MASS, and ABSTRACT. Furthermore, the data set provides insight about which pronouns participants use to refer to nouns occupying different positions on the Individuation Hierarchy (Audring 2009:124). The uniformity scores of the twenty nouns reveal high inter-speaker agreement for nouns in which semantic gender matches lexical gender (in this case masculine/feminine animates and masculine concrete count nouns) and less agreement for nouns in which semantic gender clashes with lexical gender, in particular at the right edge of the Hierarchy, i.e., where the abstract and mass nouns are located. Overall, increased variation indicates an ongoing shift in pronominalization, but for some nouns this shift is already further advanced, i.e., the most frequently chosen pronoun does not match the lexical gender of the noun. This provides evidence that semantic gender is not only found in spontaneous speech settings, but also in formal (written) language settings when speakers are more aware of (i.e., pay more attention to) their linguistic choices, and are even being assessed on them.

(iii) How are these patterns related to the gender and age of participants?

The data set also enabled us to identify groups of speakers based on their pronominalization strategies. In total, three different speaker clusters can be distinguished. Previous research has shown that the semantic system is more strongly present in the speech of younger generations and that an increase in lexical agreement is expected in the language of older speakers (e.g., Audring 2009, Bouma 2018, De Vos et al. 2021). This age-related variation also emerges in our data set, with the number of older participants gradually decreasing from cluster 1 up to cluster 3 and the number of younger participants gradually increasing from cluster 1 up to cluster 3, a case of apparent time language change. This gradual difference is also reflected in the test scores, which we used as a measure of conservatism. Cluster 1 contains the higher test scores, thus reflecting participants that adhere more closely to 'conservative' lexical agreement. These speakers mostly rely on the lexical gender of a noun when choosing a pronoun. By contrast, clusters 2 and 3 contain the (mid to) low test scores, indicating participants who are more likely to shift to innovative semantic agreement. These speakers increasingly rely on the noun's semantics when choosing a pronoun, which is reflected in the observed tendencies of masculinization (i.e., referring to non-masculine concrete count nouns with masculine pronouns) and resemanticization (i.e., referring to less individuated nouns with neuter pronouns). Our data set did not reveal convincing gender-related variation patterns (e.g., Bouma 2018, De Vos et al. 2021). There is only limited evidence that the pronominal shift is further advanced in women, as female speakers were slightly more likely to use semantic agreement than male speakers.

Comparing the three clusters based on their number of speakers, cluster 2 is the largest. This cluster could be considered a transitional one, as it displays a (partial) switch from lexical to semantic gender agreement that is also found in cluster 3. In cluster 3, the variation in pronoun assignment increases further due to the (slightly) higher proportion of feminine pronouns in reference to non-feminine nouns, especially neuter concrete countable objects. This shift is unexpected in the light of masculinization and should therefore still be accounted for. Possibly, "masculinization has opened the door to a movement in the opposite direction" here (Audring 2009:48). Audring (2009:48-51) ascribes the use of the feminine pronoun to its status as a marker of high style. As the masculine-feminine distinction is only marginally alive in (the north of) the Dutch language area, the use of the feminine pronoun can mark the speaker's command of this distinction. This may lead to the overuse of the feminine pronoun *zij* 'she' and, especially, the possessive pronoun *haar* 'her' (so-called *haar-ziekte* 'her-disease') for collectives and inanimate entities in specific registers such as journalistic and administrative writing. Audring (2009:51) argues that this use of the feminine pronoun might be detached from its usual semantic properties and that it has little to do with 'ordinary' gender agreement. Therefore, this use is not or hardly found in spontaneous speech data. Whether the increase in the use of feminine pronouns in cluster 3 is due to this type of hypercorrection cannot be determined based on the current data set. However, the method of pronoun elicitation (i.e., written assessment task) might have been a trigger to overuse the feminine pronoun if we assume it to be a written-language phenomenon symptomatic of seeking to convey an educated image.

5.2 Discussion

Although the current study was able to reveal quite robust noun and speaker patterns because of the very large sample, the methodology and subsequent data set also have some weaknesses that need to be addressed. First of all, the number of nouns included in the experiment is low. This complicates generalizing across noun classes, despite the clear distinctions that emerge between different categories in the current data set. Second, one could argue that including nouns that refer to feminine persons such as *tante* 'aunt' is too uninteresting due to their lack of variability, i.e., speakers invariably choose the feminine pronoun. However, research into southern Dutch dialects has revealed remarkable deviations in this respect. For example, it is possible for (some) Limburg speakers to refer to female persons with neuter pronouns, e.g., *het* 'it' (Piepers et al. 2021), and for (some) southern Dutch dialect speakers to refer to female persons with masculine pronouns, e.g., *hij* 'he' (Piepers et al. 2023). Furthermore, we observe adnominal gender shifts in the southern Dutch (Brabant) dialects, for example masculine suffixes on determiners preceding feminine nouns (e.g., *ene vrouw* as discussed in section 1; Doreleijers et al. 2020).

This seamlessly brings us to another limitation of the current data set, i.e. the fact that it does not contain geographical information and information about the first language of the participants, see also footnote 3. Since the platform used to disseminate the questionnaire (i.e., *Quest*) did not allow us to include questions about the regional background and linguistic background of the participants (due to privacy reasons), it was not possible to differentiate for geographical area and dialect knowledge in the analysis. This would have been very interesting, though, especially since, as discussed above, the northern part of the language area is generally considered to be more innovative and the southern part more conservative based on the gender systems of the corresponding dialects.

A final limitation relates to one of the test items, viz. the noun *onzin* 'nonsense'. The context used to elicit this item is given in (7).

(7) Jouw onzin voegt niets toe. ____ slaat nergens op.
your nonsense adds nothing to hits nowhere on
'Your nonsense adds nothing. ____ makes no sense.'
A: *Hij* 'he'
B: *Zij* 'she'

C: *Het* 'it'

In choosing an appropriate pronoun in this context, participants may have selected one that refers back to the whole first sentence rather than to the noun *onzin* 'nonsense' specifically. This might also explain why this noun behaves differently from the other abstract nouns in cluster 1. Speakers in this cluster in general adhere to the lexical agreement system, except for this noun, which is assigned a neuter pronoun.

In spite of the above-mentioned limitations, the current study adds a new perspective to existing research on lexical versus semantic agreement. It shows that processes of masculinization and resemanticization not only occur in spontaneous speech settings, but that they also emerge in formal assessment contexts that yield a high level of salience and linguistic awareness. The current data set is consistent with the research of Audring (2009), that full pronouns, even in these more formal settings, also have a tendency towards semantic agreement. By continuing to include new data sets of speakers from different generations into the debate, we are hopefully on track to unravel the many mysteries of pronominal gender shift in Dutch.

References

Audring, Jenny. 2006. Pronominal gender in spoken Dutch. *Journal of Germanic Linguistics* 18:85–116.

Audring, Jenny. 2009. Reinventing pronoun gender. Doctoral Dissertation, VU Amsterdam.

Audring, Jenny. 2021. Veelzijdig onzijdig. In *Wat gebeurt er in het Nederlands?! Over taal, frequentie en variatie*, ed. Nicoline van der Sijs, Lauren Fonteyn, and Marten van der Meulen, 79–84. Gorredijk: Sterck & De Vreese.

Audring, Jenny, and Geert Booij. 2009. Genus als probleemcategorie. In *Perspectieven op genus in het Nederlands. taal en tongval. themanummer 22*, ed. Leonie Cornips and Gunther De Vogelaer, 13–37.

de Bont, Anton P. 1962. *Dialect van Kempenland; meer in het bijzonder D'Oerse taol. Deel I: Klank- en vormleer en enige syntaktische bijzonderheden*. Assen: Van Gorcum.

Bouma, Gosse. 2018. Agreement mismatches in Dutch relatives. *Belgian Journal of Linguistics* 31:137–164.

Brouwer, Susanne, Simone Sprenger, and Sharon Unsworth. 2017. Processing grammatical gender in Dutch: evidence from eye movements. *Journal of Experimental Child Psychology* 159:50–65.

Cardinaletti, Anna, and Michal Starke. 1999. The typology of structural deficiency: a case study of three classes of pronouns. In *Clitics in the languages of Europe*, ed. Henk van Riemsdijk, 145–233. Berlin: Mouton.

Cohen, Jacob. 1962. The statistical power of abnormal-social psychological research. *Journal of Abnormal and Social Psychology* 65:145–153.

Cornips, Leonie, and Gunther De Vogelaer. 2009. Variatie en verandering in het Nederlandser genus: een multidisciplinair perspectief. *Taal & Tongval* 61:1–12.

Cornips, Leonie, and Aafke Hulk. 2008. Factors of success and failure in the acquisition of grammatical gender in Dutch. *Second Language Research* 24:267–295.

De Paepe, Jessie, and Gunther De Vogelaer. 2008. Grammaticaal genus en pronominale verwijzing bij kinderen. Ee taalverwervingsperspectief op een eeuwenoud grammaticaal probleem. *neerlandistiek.nl* 08.02:1–23.

De Troij, Robbert, Stef Grondelaers, and Dirk Speelman. 2023. Natiolectal variation in dutch morphosyntax: A large-scale, data-driven perspective. *Journal of Germanic Linguistics* 1:1–68.

De Vogelaer, Gunther. 2009. Changing pronominal gender in Dutch: transmission or diffusion? In *Language variation, European perspectives ii*, ed. Stavroula Tsiplakou, M. Karyolemou, and P. Pavlou, 71–80. Amsterdam: John Benjamins.

De Vogelaer, Gunther, and Gert De Sutter. 2011. The geography of gender change: pronominal and adnominal gender in Flemish dialects of Dutch. *Language Sciences* 33:192–205.

De Vos, Lien. 2009. De dynamiek van hersemantisering. *Taal & Tongval* 61:82–110.

De Vos, Lien, and Gunther De Vogelaer. 2011. Dutch gender and the locus of morphological regularization. *Folia Linguistica* 45:245–281.

De Vos, Lien, Gunther De Vogelaer, and Gert De Sutter. 2021. Weighing psycholinguistics and social factors for semantic agreement in Dutch pronouns. *Journal of Germanic Linguistics* 33:30–66.

Doreleijers, Kristel, Marjo van Koppen, and Jos Swanenberg. 2020. De dynamiek van geslachtsmarkering in de Noord-Brabantse dialecten. *Taal & Tongval* 72:69–116.

Doreleijers, Kristel, Joske Piepers, Ad Backus, and Jos Swanenberg. 2021. Language variation in dialect-standard contact situations. *Applications of Cognitive Linguistics* 175.

Geeraerts, Dirk. 1992. Pronominale masculiniseringsparameters in Vlaanderen. In *De binnenbouw van het nederlands: een bundel artikelen voor Piet Paardekooper*, ed. Hans Bennis and Jan de Vries, 73–84. Dordrecht: ICG Publications.

Geeraerts, Dirk, Stefan Grondelaers, and Dirk Speelman. 1999. *Convergentie en divergentie in de nederlandse woordenschat. een onderzoek naar kleding- en voetbal namen*. Amsterdam: Meertens Institute.

Geerts, Guido. 1966. *Genus en geslacht in de Gouden Eeuw: Een bijdrage tot de studie van de nominal klassifikatie en daarmee samenhangende adnominale flexievormen en pronominale verschijnselen in Hollands taalgebruik van de zeventiende eeuw*. Brussel: Belgisch Interuniversitair Centrum voor Neederlandistiek.

Greenacre, Michael. 2007. *Correspondence analysis in practice*. London & New York: Chapman & Hall, 2nd edition.

Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jaap de Rooij, and Maarten Cornelis van den Toorn. 2019. *Algemene Nederlandse Spraakkunst (e-ans)*.

Hinskens, Frans, Roeland van Hout, Pieter Muysken, and Ariën van Wijngaarden. 2021. Variation and change in grammatical gender marking: the case of dutch ethnolects. *Linguistics* 1:75–100.

Hoppenbrouwers, Cor. 1983. Het genus in een Brabants regiolect. *Tabu* 13:1–25.

Hoppenbrouwers, Cor. 1990. *Het regiolect: van dialect tot Algemeen Nederlands*. Muiderberg: Coutinho.

Husson, Francois, Julie Josse, Sebastien Le, and Jeremy Mazet. 2014. *Factominer: Multivariate exploratory data analysis and data mining with r*. URL `http://CRAN.R-project.org/package=FactoMineR`, R package version 1.26.

Husson, Francois, Julie Josse, and Jérôme Pagès. 2010. Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing data? Technical Report - Agrocampus.

Husson, François, Sébastien Lê, and Jérôme Pagès. 2011. *Exploratory multivariate analysis by example using R*. Boca Raton/London/New York: CRC Press.

Josefsson, Gunlög. 2006. Semantic and grammatical genders in swedish: independent but interacting dimensions. *Lingua* 116:1346–1368.

Kraaikamp, Margot. 2012. The semantics of the Dutch gender system. *Journal of Germanic Linguistics* 24:193–232.

Kraaikamp, Margot. 2017. Semantic versus lexical gender. Synchronic and diachronic variation in Germanic gender agreement. Doctoral Dissertation, University of Amsterdam.

Labov, William. 1972. The isolation of contextual styles. In *Sociolinguistic patterns*, 70–109. Philadelphia: University of Pennsylvania Press.

Labov, William. 1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2:205–254.

Levshina, Natalia. 2015. *How to do linguistics with R. Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Mills, Anne E. 1986. *The acquisition of gender. A study of English and German*. Berlin: Springer.

Piepers, Joske, Ad Backus, and Jos Swanenberg. 2021. *Ziej* is a woman and *het* is a girl: A referent's age guides pronominal gender variation in Limburgian. *Taal & Tongval* 73:1–44.

Piepers, Joske, Jos Swanenberg, and Ad Backus. 2023. 'he' is still here: On the contemporary use of masculine pronouns for women in Dutch dialects. Ms. Tilburg University.

R Core Team. 2014. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`.

Richardson, John T.E. 2011. Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review* 6:135–147.

Romijn, Kirsten. 1996. *Hoe doen we het? Verwijzen naar linguïstische en cognitieve representaties met het voornaamwoord 'het'*. Amsterdam: P.J. Meertens Instituut.

Sasse, Hans-Jürgen. 1993. Syntactic categories and subcategories. In *Syntax. Ein internationales Handbuch zeitgenössisscher Forschung. An international handbook of contemporary research*, ed. Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, and Theo Venneman, 646–686. Berlin: De Gruyter.

Siemund, Peter. 2008. *Pronominal gender in English: a study of English varieties from a cross-linguistic perspective*. London: Routledge.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In *Grammatical categories in Australian languages*, ed. Richard Dixon, 112–171. Canberra: Australian Institute of Aboriginal Studies.

Stroop, Jan. 1989. Woordgeslacht ('genus'). *De Vierschaer* 7:4–15.

Weijnen, Antonius Angelus. 1971. *Schets van de geschiedenis van de Nederlandse syntaxis*. Assen: Van Gorcum & Comp.

Wurzel, Wolfgang Ullrich. 1986. Die wiederholte Klassifikation von Substantiven. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 39:76–96.

Appendix: the test items used in the questionnaire

(1) Als het regent draag ik een hoed. ____ hangt nu aan de kapstok.
if it rains wear I a hat hangs nu at the coat.rack
'When it rains, I wear a hat. ____ now hangs on the coat rack.'

(2) Mijn tante is erg ongeduldig. ____ ijsbeert door de kamer.
my aunt is very impatient. paces through the room.
'My aunt is very impatient. ____ paces around the room.'

(3) In de straat staat een auto. ____ is van de buren.
in the street stands a car is of the neighbors
'There is a car in the street. ____ belongs to the neighbors.'

(4) Achter het huis ligt een tuin. ____ staat vol met bloemen.
behind the house lies a garden stands full with flowers
'Behind the house is a garden. ____ is full of flowers.'

(5) Ik heb een hemd gestreken. ____ zat vol lelijke vouwen.
I have a shirt ironed sat full ugly creases
'I ironed a shirt. ____ was full of ugly creases.'

(6) Ik ken alle informatie al. ____ is mij bekend.
I know all information already is me known
'I already know all the information. ____ is familiar to me.'

(7) Mijn bed staat in de kamer. ____ is warm en zacht.
my bed stands in the room is warm and soft
'My bed is in the room. ____ is warm and soft.'

(8) Het dorp heeft een bibliotheek. ____ is vandaag geopend.
the village has a library is today opened
'The village has a library. ____ was opened today.'

(9) Mijn boek ligt hier. ____ is groot en dik.
my book lies here is big and thick
'My book lies here. ____ is big and thick.'

(10) Er komt een boer op televisie. ____ verbouwt de boerderij.
there comes a farmer on television renovates the farm
'A farmer appears on television. ____ is renovating the farm.'

(11) Wat een hitte! ____ droogt de natuur uit.
what a heat dries the nature out
'What a heat! ____ is drying out nature.'

(12) Mijn oma heeft een taart gebakken. ____ is jarig.
my grandma has a cake baked is having.her.birthday
'My grandma baked a cake. ____ is celebrating her birthday.'

(13) Jouw onzin voegt niets toe. ____ slaat nergens op.
your nonsense adds nothing to hits nowhere on
'Your nonsense adds nothing. ____ makes no sense.'

(14) Mijn koekje is al op. ____ was erg lekker.
my cookie is already up was very tasty
'My cookie is already finished. ____ was very tasty.

(15) De buren hebben een hond. ____ rent door de tuin.
the neighbors have a dog runs through the garden
'The neighbors have a dog. ____ runs through the garden.'

(16) In het park staat een schommel. ____ staat in de speeltuin.
in the park stands a swing stands in the playground
'There is a swing in the park. ____ is in the playground.'

(17) Gisteren was er nog honing. Vandaag is ____ op.
yesterday was there still honey today is up
'Yesterday there was honey. Today ____ has run out.'

(18) Een koe geeft goede melk. ____ wordt dagelijks gemolken.
a cow gives good milk becomes daily milked
'A cow produces good milk. ____ is milked daily.'

(19) In de verte zie ik een vliegtuig. ____ gaat bijna opstijgen.
in the distance see I a plane goes almost take.off
'In the distance, I see a plane. ____ is about to take off.'

(20) Ons paard is erg sierlijk. ____ wint alle dressuurwedstrijden.
our horse is very graceful wins all dressage.competitions
'Our horse is very graceful. ____ wins all dressage competitions.'