

## Weighing Psycholinguistic and Social Factors for Semantic Agreement in Dutch Pronouns

Lien De Vos

*Université de Liège*

Gert De Sutter

*Ghent University*

Gunther De Vogelaer

*University of Münster*

Previous research has shown that Dutch pronominal gender is in a process of resemanticization: Highly individuated nouns are increasingly referred to with masculine and feminine pronouns, and lowly individuated ones with the neuter pronoun *het*/'t 'it', irrespective of the grammatical gender of the noun (Audring 2009). The process is commonly attributed to the loss of adnominal gender agreement, which is increasingly blurring distinctions between masculine and feminine nouns and, therefore, requires speakers to resort to semantic default strategies (De Vogelaer & De Sutter 2011). Several factors have been identified that influence the choice of semantic vis-à-vis lexical agreement, both linguistic and social. This article seeks to weigh the importance of both structural and social factors in pronominal gender agreement in Belgian Dutch, using the Belgian part of the Spoken Dutch Corpus. A multivariate statistical analysis reveals that most effects are structural, including noun semantics and the syntactic function of the antecedent and the pronoun, as well as the pragmatic status of the antecedent. The most important social factor is speech register. We argue that these effects support a psycholinguistic account in which resemanticization is seen as a change from below, caused by hampered lexical access to noun gender.

**Keywords:** grammatical gender, Dutch, pronouns, lexical agreement, semantic agreement

### 1. Introduction.

Previous research has shown that agreement between pronouns and their antecedent nouns in Dutch is increasingly shifting from a lexical to a semantic system of reference: In the former type of pronominal reference, the pronoun's form is determined by the lexical gender of the

© Society for Germanic Linguistics 2021. This is an Open Access article, distributed under the terms of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

antecedent noun, whereas in the latter system, the pronoun's form is determined by the noun's semantics. One of the main determinants of this process, which is referred to as a RESEMANTICIZATION process following Audring (2006), is the degree of individuation of the antecedent noun: Highly individuated nouns are increasingly referred to with masculine and feminine pronouns, and lowly individuated ones with neuter *het/'t* 'it', irrespective of their grammatical gender. On the high end of the Individuation Hierarchy (Sasse 1993, Siemund 2008), in the animate domain, present-day Dutch more commonly applies the so-called natural gender rule than older stages of the language (or the closely related German; Kraaikamp 2017:63–73). Thus, a neuter noun such as Dutch *meisje* 'girl' is more often pronominalized with *zij/ze* 'she' instead of *het/'t* 'it' than its neuter German equivalent *Mädchen* with *sie* 'she' instead of *es* 'it'. The same generalization holds for other neuter nouns referring to entities that carry biological gender. This article is concerned with the inanimate end of the Individuation Hierarchy, however, as illustrated in 1, in which a mass noun, namely, *wijn* 'wine', refers to an unspecific quantity. In these situations, Dutch shows an increasing tendency to use the neuter pronoun (in this case the reduced form of *het*, namely, *'t*), even if the noun has common gender (or is masculine in three-gender varieties of Dutch). The common gender is indicated here and below as C and the neuter gender as N.

- (1) een decanteerfles. daar stop je je wijn in  
 a decanter(C) there put you your wine(C) in  
 en dan kan 't luchten.  
 and then can 3.N breathe

'A decanter; You put your wine in it and then it can breathe.'

(Audring 2009:97)

About a decade of research, where methods have been used ranging from corpus studies via questionnaires to experiments, has shown the resemanticization process to be highly complex, in that a substantial number of both linguistic and social factors were found to be relevant. There have been few attempts to weigh the different factors at play, however. This article, therefore, aims to provide a more comprehensive overview of the factors influencing the choice between lexical agreement

and semantic agreement, using multivariate statistics to weigh their importance. In addition to being a highly complex process, there appears to be some geographical variation in the way in which resemanticization is implemented in the Netherlands (especially in the Holland region; Audring 2009) and in Belgium (De Vos & De Vogelaer 2011). Since corpus investigations (especially Audring 2009, Bouma 2018) so far have focused on the Netherlands, this article investigates Belgian corpus data, basically complementing Audring's (2009) analysis of the northern part of the *Corpus Gesproken Nederlands* (Spoken Dutch Corpus).

This paper is structured as follows: Section 2 reviews the literature on resemanticization in Dutch, section 3 presents the corpus materials and the annotation procedure, and in section 4 and 5 the results of the study are presented and discussed. Finally, section 6 summarizes the main findings and formulates predictions to be tested in future research.

## **2. Resemanticization: A Change in Progress.**

### *2.1. Background: Change in the Dutch Gender System.*

Even though a certain proportion of semantic agreement, including the use of neuter pronouns for lowly individuated entities, is likely inherited from older stages of Dutch (Kraaikamp 2017), a lot of the present-day variation reflects change in progress. Indeed, historically speaking, Dutch seems to be in a transition from a system in which pronouns show lexical agreement with their antecedent, as found in German, to a system in which the semantics of the referent is more important, as in English. Since gender systems all tend to have a semantic core, and semantics is, in fact, assumed to have played a role in the very origin of the Indo-European gender system (Luraghi 2011), this development is dubbed resemanticization by Audring (2006), who adopts a term from Wurzel (1986). While it is still unknown when exactly each of the stages in this process took place, different explanations of changing pronominal gender have been proposed (Geerts 1966, De Vogelaer & De Sutter 2011, Kraaikamp 2012). These explanations converge on the idea that a number of developments in the Modern Dutch period (starting around 1500 C.E.) weakened the grammatical gender system. As a result, grammatical gender became a problematic category for present-day users, which, in turn, lead to the resemanticization process (Audring & Booij 2009). First, in contrast with German, gender assignment is believed to have become largely arbitrary in Dutch: While many formal

and semantic assignment regularities were detected for German (for example, Köpcke & Zubin 1983), these appear to be absent in Dutch (Durieux et al. 1999). Second, many of the gender distinctions upheld in older stages of the language have become blurred. Gender distinctions between masculine and feminine gender have become particularly weak, to the extent that they have collapsed into the common gender in northern varieties, including present-day Standard Dutch. In colloquial varieties spoken in the southern part of the language area, most notably in Belgium, the masculine/feminine distinction shows the effects of syncretism and deletion of inflectional endings. Table 1 shows a typical gender paradigm in three varieties of Dutch, with agreement patterns of articles and adjectives in definite and indefinite NPs.<sup>1</sup> While both southern varieties still distinguish three genders, the final *-n* marking masculine gender is typically dropped before all consonants except *b/d/t/h*, thereby causing the masculine/feminine distinction to be blurred. The difference between masculine/feminine gender and neuter gender is somewhat more stable, especially as the neuter definite article *het/'t* ‘the’ is not exposed to processes of schwa-deletion and *n*-deletion that have affected large parts of the morphology of Dutch.

	Masculine ‘a/the beautiful chair’	Feminine ‘a/the beautiful table’	Neuter ‘a/the beautiful sink’
Northern: Standard Dutch	<b>een</b> mooi- <b>e</b> stoel/tafel <b>de</b> mooi- <b>e</b> stoel/tafel		<b>een</b> mooi aanrecht <b>het</b> mooi- <b>e</b> aanrecht
Southern: West Flemish	<b>e</b> mooi- <b>e(n)</b> stoel <b>de(n)</b> mooi- <b>e(n)</b> stoel	<b>e</b> mooi- <b>e</b> tafel <b>de</b> mooi- <b>e</b> tafel	<b>e</b> mooi aanrecht <b>het</b> mooi aanrecht
Southern: East Flemish	<b>ne(n)</b> mooi- <b>e(n)</b> stoel <b>de(n)</b> mooi- <b>e(n)</b> stoel	<b>een</b> mooi- <b>e</b> tafel <b>de</b> mooi- <b>e</b> tafel	<b>e(en)</b> mooi aanrecht <b>het</b> mooi aanrecht

Table 1. Gender marking in three varieties of Dutch.

Apart from gender syncretism, extensive dialect loss is observed in large parts of the Dutch language area, which renders the distinctions

<sup>1</sup> Apart from these agreement targets, demonstratives and relative pronouns show patterns of syncretism. See De Schutter et al. 2005 for a comprehensive overview of gender agreement in different varieties of Dutch.

between masculine and feminine nouns even less visible (see Plevoets et al. 2009 for figures). As a result, the learnability of the pronominal system is affected, both in northern and in southern varieties (see Audring 2014 and De Vos & De Vogelaer 2011, respectively). In particular, speakers increasingly resort to semantic default strategies in pronominal reference, which, at least for inanimate referents, results in an increased usage of the neuter pronoun *het*/'it', as exemplified in 1 above (De Vogelaer & De Sutter 2011).<sup>2</sup> A scenario involving learnability yields a number of hypotheses regarding linguistic and social factors influencing the choice in favor of semantic vis-à-vis lexical agreement in varieties in which the change is in progress. Some of these factors have been described in earlier work on resemanticization in Dutch, whereas others can be derived from more general work on lexical versus semantic agreement, the use of referential expressions, or language change.

## 2.2. Linguistic Factors.

Linguistically speaking, an account of resemanticization involving the learnability of the gender system predicts lexical agreement to be more likely for nouns whose gender is more easily learned. This includes nouns with morphologically marked gender, in particular with derivational suffixes marking feminine gender, such as *-ing* in *herstelling* 'repairment' or *-de/-te* in *grootte* 'size', which have been hypothesized to inhibit change (Haeseryn et al. 1997, §3.3.3.4). In a similar vein, highly frequent nouns are found to trigger more lexical agreement in southern varieties of Dutch (De Vos & De Vogelaer 2011). More importantly, a noun's propensity for lexical vis-à-vis semantic agreement is linked to its position in the Individuation Hierarchy (Sasse 1993:659, Siemund 2008:140, Audring 2009:124), a variant of the Animacy Hierarchy

---

<sup>2</sup> There is disagreement as to which development triggered resemanticization. Audring (2009:27–29) suggests that lexical agreement in pronouns is unstable because the tryadic pronominal paradigm distinguishes masculine, feminine, and neuter pronouns, which would not match with a binary distinction between common and neuter gender in the nominal domain. In this article, it is shown that an unstable masculine/feminine distinction can also trigger resemanticization, without a collapse of both genders (see also De Vogelaer & De Sutter 2011).

(Silverstein 1976), which is widely applied in typological research.<sup>3</sup> The Individuation Hierarchy ranks nouns in terms of the degree of individuation of their respective referents. Referents are most highly individuated when they are human (for example, unique persons such as *Peter* or more general references such as *man* or *nurse*). Next in line are other animates (a category including animals), which are more individuated than bounded objects, which are more individuated than abstract nouns. At the rightmost position are mass nouns. Semantically speaking, referents at the left end of the hierarchy are more compatible with masculine/feminine pronouns, whereas the rightmost referents, which are investigated in this article, tend to be more compatible with neuter pronouns.

## (2) The Individuation Hierarchy (Audring 2009:124)

Human > Other > Bounded object/ > Specific > Unspecific mass,  
                   animate    abstract                    mass            unbounded abstract

The Individuation Hierarchy resists decomposition into binary categories because many of the properties that underlie these categories (such as being human, animateness, boundedness, abstractness, etc.) are not necessarily mutually exclusive and are not always conceived of as discrete. This has been illustrated for resemanticization in Dutch by Fletcher (1987) and Verhoeven (1990:502). In addition, individuation cannot be reduced to inherent properties of referents, as it is sensitive to contextual factors, including the degree to which speakers construe a referent as being unique or bounded (see Audring 2009 and Bouma 2018). For instance, in the example in 3 from the Spoken Dutch Corpus, the noun *aandacht* ‘attention’ is used as a count noun in a highly individuated way, in that it can be divided and thus appears bounded, which illustrates that, especially in a given context, classifying nouns as

---

<sup>3</sup> More precisely, semantic agreement becomes more likely as the conflict between lexical and semantic agreement intensifies, that is, in particular at the ends of the Individuation Hierarchy. An example of a highly individuated noun would be *meisje* ‘little girl’, which has neuter gender but is often referred to with feminine pronouns. This article does not discuss highly individuated nouns, namely, animates and especially humans, so within the scope of this paper, semantic gender is especially to be expected for lowly individuated entities.

being count/mass or concrete/abstract is far from straightforward.

- (3) maar ge moet ge moet hem splitsen hè?  
 but you have.to you have.to 3.M divide right  
 uwen aandacht.  
 2.POSS-M attention(M)

‘But you have to, you have to divide it, right? Your attention.’

The methodological challenge of operationalizing individuation has been addressed in different ways in studies on resemanticization in Dutch. Audring (2009:127), for instance, tests the relevance of the Individuation Hierarchy as given in 2, but assumes a cut-off point between the domain of masculine vis-à-vis neuter pronouns within the category of specific mass. This is not clearly defined but operationalized in terms of binary oppositions between count versus mass and concrete versus abstract nouns (Audring 2009:159). Similarly, De Vos & De Vogelaer’s (2011) questionnaire study works with test sentences containing prototypical instances of four categories defined by the features *concrete/abstract* and *count/mass*, and finds concrete count nouns to show some 10% more lexical agreement than the other categories, which behave rather similarly.

The data from De Vos & De Vogelaer (2011) come from the same region as the corpus data for this study, so they provide an empirical argument in favor of a more straightforward, binary distinction between concrete, tangible objects that have “spatial integrity” (Josefsson 2006: 1352) on the one hand, and entities lacking such spatial integrity on the other. The latter can be conceived as heterogeneous collections, semantic networks or abstractions (Romijn 1996:40–42). Following Romijn’s (1996:37–42) study of the use of neuter pronouns in Dutch, the former are referred to as “simple” entities in this study, whereas the latter are referred to as “complex”. Adopting this distinction implies that nouns such as in 3, which are included in categories with an intermediate ranking on the Individuation Hierarchy (such as bounded abstract and specific mass in Audring 2009) are considered semantically complex in this study.

In addition to the semantics of the referent, there are contextual factors that may influence the choice of pronouns as referring expressions, which can be organized into three categories, namely, i) the

status of the referent in discourse, ii) the context in which the anaphoric pronoun appears, and iii) the distance between antecedent and pronoun. We consider each of these three sets below.

The discourse status of the referent has been studied under a variety of names and theories, including *focusing* (Ward 1985), *givenness* (Gundel et al. 1993), *topicality* (Givón 1983), and *accessibility* (Ariel 1990) in pragmatics, or *prominence* (Gordon & Hendrick 1998) and *salience* (Arnold 1998) in the literature on cognitive structure and processing efficiency. These notions only partially overlap, but, crucially, what they have in common is being highly multifactorial in nature and resisting implementation in terms of a small number of binary features (see, for instance, Siewierska 2004:174–210 for a discussion of *accessibility*, and Rose 2005 on *salience*). Most of these cognitive models connect the discursive status of the referent, that is, its degree of givenness, accessibility, etc. to the choice of referring expressions, with pronouns in general ranking high on different scales.

For the most part, the relevance of such hierarchies for the alternation between lexical and semantic agreement in Dutch has not been considered closely so far (for example, by Fletcher 1987 and Semplicini 2012). Hence, they were included in the investigation, in two ways. First, the influence of the antecedent determiner on the choice between lexical and semantic agreement was investigated. Second, the role of the antecedent's grammatical function was also analyzed. As, for instance, captured in Keenan & Comrie's 1977 Accessibility Hierarchy, grammatical function can indeed be considered a main determinant of salience, and as such it potentially influences the availability of pronominalization strategies and agreement options, among many other things (for instance, relativization in Keenan & Comrie 1977).

The second discourse factor concerns the context in which the anaphoric pronoun appears. Indeed, the degree to which arguments are construed or can be perceived as individuated also relates to the predicate a pronoun combines with, its thematic role and syntactic function, which may influence the choice in favor of lexical vis-à-vis semantic agreement. Semantic parameters relating to the predicate have been suggested for pronominal gender in historical varieties of English (for instance, Grund 2011), but they have hardly been explored for Dutch, apart from Romijn's (1996:38) account. Romijn suggests that semantic *het/'t* 'it' is more likely to be found in copular clauses, as in 4a, than in



noncopular ones, as in 4b,c (the examples are from Romijn 1996:38)

(4) Q: Heb je die paarse bril met groene pootjes gezien?  
 have you DEF.C purple.C glass(C) with green arms seen  
 ‘Did you see those purple glasses with green arms?’

A: a. Ja het/hij is vreemd.  
 yes it/he is weird  
 ‘Yes, they are weird.’

b. Ja, ik heb het/’m vorige week zelfs gepast.  
 yes I have it/him last week even fitted  
 ‘Yes, I even tried them on last week.’

c. Ja, het/hij lijkt een beetje op die andere met roze pootjes.  
 yes it/he looks a bit on that other with pink arms  
 ‘Yes, they look a bit like the other ones, with pink arms.’

In noncopular clauses, Romijn (1996) also discusses differences between subjects versus objects (4b versus 4c), as does Audring (2009:162–164), who finds a higher propensity for semantic agreement in subject pronouns. Audring explains this contrast in terms of case marking (that is, nominative pronouns show a stronger tendency toward semantic agreement) and not syntactic function, since the effect is not found with demonstrative pronouns, which are no longer marked for case in Dutch.

The third discourse factor, namely, the distance between antecedent and pronoun, has also already been investigated in previous studies on semantic agreement in Dutch. In line with the Agreement Hierarchy formulated on the basis of typological research (Corbett 1979), Audring (2009:164–167) shows that a considerable interval between the two indeed increases the likelihood of semantic agreement.

In addition to the semantics of the referent and the three discourse factors mentioned above, a final linguistic factor is the distinction between full and reduced pronouns, which is found in all three genders. Thus, Standard Dutch has the full pronoun *hij* and its reduced version *-ie* in masculine gender, *zij* versus *ze* in feminine gender, and *het* versus *’t* in neuter gender. These pronouns constitute the vast majority of corpus examples, even though some additional, regional forms are included as well. The distinction between full and reduced pronouns determines the

referential potential of the forms, in that most full forms are only used in reference to humans (Haeseryn et al. 1997 §5.2.7). While pronouns referring to humans fall outside the scope of the investigation, this factor was included as well, even though its relevance for resemanticization is unclear. The underlying prediction of many salience/hierarchy approaches is that full pronouns are used for less salient antecedents, whereas reduced pronouns are used for more salient ones (see Ariel 1990). Under this view, full pronouns are good candidates for showing semantic agreement, which is exploited in processes of personification (see Mills 1986, chapter 5 for a case study). Other factors may be involved, however, such as the tendency to overuse full forms in formal registers, which Audring (2009:155) considers the reason for them to be more conservative, that is, show more lexical agreement. An additional consideration is that the paradigm of reduced forms contains gaps. Most prominently, many varieties of Dutch, including Standard Dutch, only use the reduced masculine subject form *-ie* in enclitic positions, that is, attached to a complementizer or a verb. While the effect of this on agreement relations is hard to predict, the distinction between full and reduced forms is included in the analysis.

### 2.3. *Social Factors.*

In addition to the linguistic factors mentioned above, the proposed process of resemanticization also implies that certain social factors are relevant. First, in general, change in progress is typically implemented by successive generations of speakers pushing a phenomenon beyond levels observed in previous generations. Second, changes tend to be, on average, further advanced in women (Labov 1990). In case of the shift from lexical to semantic agreement, age appears to be a significant factor as well. Audring (2009:169) finds a significantly stronger tendency to use semantic agreement in younger speakers in the Dutch part of the Spoken Dutch Corpus: Speakers below the age of 20 use semantic agreement over 70% of the time versus ca. 40% found among speakers aged 60 and above. However, some intermediate age groups show an unexpectedly high or low proportion of semantic agreement. Bouma's (2018) study of semantically agreeing relative pronouns on Twitter also shows age differences, but the steep increase in the use of semantic agreement in the age groups born after 2000 (in comparison to the age group born in the 1980s) may relate to genre-specific differences

between the writing of so-called digital natives and older age groups (see Verheijen 2018 for illustrations from Dutch). Finally, for Belgium, some questionnaire studies (De Vogelaer & De Sutter 2011, De Vos & De Vogelaer 2011) have also revealed significant differences between age groups, but with small effect sizes. None of these studies report any gender differences.

The potential effects of three additional social factors are less clear, namely, geography, social class, and speech register. Statistically speaking, semantic agreement is particularly common in northern varieties of Dutch, including Standard Dutch as spoken in the Netherlands, which functions as the prestige variety in the Dutch language area. However, the degree to which different regions in Belgium orient themselves toward Standard Dutch does not seem to correlate with resemanticization, since the phenomenon is especially widely attested in the province of West Flanders, which more strongly resists standardization processes than other regions (De Vogelaer & De Sutter 2011).

In addition, there are no indications that semantic agreement is linked to specific social classes of speakers (as typically defined via educational levels or income classes), or to certain registers.<sup>4</sup> For instance, most reference grammars discuss geographical, social and/or stylistic connotations associated with the use of masculine vis-à-vis feminine pronouns when referring to non-neuter inanimate nouns but do not treat *het*/*'t* 'it' as an alternative option in such cases (see Haeseryn et al. 1997, §3.3.3.6). Note that the website of the *Taalunie*, the Netherlands and Belgium's official body responsible for policies on the Dutch language, provides advice on 18 grammatical issues related to gender, but only one concerns resemanticization (of relative pronouns).<sup>5</sup> In

---

<sup>4</sup> As a result, the social meaning of semantic agreement contrasts strongly with that of the agreement endings marking the masculine/feminine distinction (see table 1). The latter are associated with southern Dutch and are considered sociolinguistic markers of nonstandard language usage (see Hoppenbrouwers 1983 and Cornips & Hulk 2006 on the Netherlands, and Plevvoets et al. 2009 on Belgium).

<sup>5</sup> On a more circumstantial note, it is striking that some of the key publications on semantic agreement are written by non-native speakers of Dutch, like Fletcher (1987), Siemund (2002), and Audring (2006). Historical examples of the phenomenon, dating back as far as the 17th century, only appear to have been detected later (see, among others, Van der Horst 2008, Semplicini 2012,

Labovian terms, semantically motivated *het*/'t 'it' for non-neuter nouns seems to qualify as an indicator rather than as a marker, that is, a linguistic variable more often found in certain social groups (younger speakers, women), without speakers being aware of using it. There is no particular overt or covert prestige associated with semantic gender. Hence, the resemanticization process would be an instance of change from below (the level of social consciousness) or transmission (see De Vogelaer 2009 for further argumentation).

### 3. Method: A Corpus Study.

This article aims to weigh the importance of structural vis-à-vis social factors in pronominal gender agreement in southern Dutch, using the Belgian component of the Spoken Dutch Corpus, which yields a sample of some 3 million words (Oostdijk 2000, van Eerten 2007). In its design, the study goes beyond a replication of Audring 2009 by i) integrating additional factors, especially syntactic and discursive ones, that influence the choice of semantic versus lexical agreement; and ii) providing a multivariate statistical analysis, which, in turn, allows one to distinguish between significant and nonsignificant factors and yields a more precise estimation of effect sizes. Section 3.1 describes how relevant tokens were gathered from the corpus, whereas the next one (3.2) provides more information on the factors that were analyzed.

#### 3.1. Sampling the Cases.

Since variation and change in Dutch are especially noticeable in agreement relations that involve inanimate referents, the study focuses on pronouns that refer to inanimates and leaves out of consideration all pronouns referring to humans and animals (which, in fact, represent the vast majority of tokens in the corpus). In addition, the study only deals with personal pronouns, since the personal pronoun paradigm is the only component of Standard Dutch grammar where the masculine/feminine distinction is upheld, and also the component from which innovations in reference strategies are expected to spread to other agreement targets (Corbett 1991:242; see Audring 2009:159–164 for an illustration from Netherlandic Dutch).

Having gathered all relevant pronouns from the Belgian part of the

---

and Kraaikamp 2017).

Spoken Dutch Corpus, we found that neuter nouns are almost invariably referred to by neuter pronouns (in 1,708/1,730 cases, or 98,7%). This constitutes a sharp contrast with Audring's (2009) data from the (center of) the Netherlands, where neutral nouns frequently trigger masculine or common pronominal agreement. For our Belgian case study, this renders further analysis of agreement relations with neuter nouns superfluous, so these were removed from the database.

As for the references linked to non-neuter nouns, not all pronouns in the corpus were suitable for inclusion. A decisive criterion was that an agreement relation between the pronoun and an antecedent should be established. Thus, both deictically used pronouns and pronouns lacking an overt antecedent were not analyzed because no unambiguous information can be obtained about the grammatical gender of the antecedent, which is one of the main factors in the investigation.

Further problems arise with neuter pronouns, since they have a variety of possible uses. For example, they not only refer to neuter nouns but also to entire propositions and to other types of nonentities (Kraaikamp 2012), which fall out of the scope of our analysis. In a similar vein, instances of *het/'t* as a subject of a nominal predicate of which the nominal part is introduced by a determiner (as in 5a) were not considered to establish an agreement relation, as the pattern is also possible with nouns referring to humans (see 5b; compare Haeseryn et al. 1997, §5.2.5.2.2).

(5) a. Het is mijn huis.  
3.N-SG is my house(N).  
'It is my house.'

b. Wat vind je van Dick? — Het is een aardige jongen.  
what find you of Dick 3.N-SG is a nice boy  
'What do you think of Dick?—It's a nice guy.'

Another problematic case mentioned by Haeseryn et al. (1997) are copular clauses with an adjectival predicate, as illustrated in 6a. While one cannot exclude the possibility that *het* 'it' in 6a refers to the preceding sentence, the more logical interpretation is that it refers to *jam* 'jam'. Haeseryn et al. (1997, §5.2.5.2.2) conclude that the kind of reference displayed in 6a is unclear and cannot be described. Examples such as 6b, which also contain a neuter pronoun referring to a non-neuter noun, are

considered ungrammatical by Haeseryn et al. (1997, §5.2.5.2.2).

- (6) a. Ze trakteerde op eigengemaakte jam. Het was erg lekker.  
 she treated to homemade jam(F). 3.N-SG was very tasty.  
 ‘She treated us to homemade jam. It was very tasty.’
- b. Hoe kwam ze aan die jam? —  
 how came she on DEF.C jam(F)?
- Ze had het/dat zelf gemaakt.  
 she had 3.N-SG/DEM.N herself made.
- ‘How did she get that jam? She had made it herself.’

However, in the Spoken Dutch Corpus, examples of both types are amply attested. Since this study acknowledges that pronoun usage in contemporary Dutch is not always in line with the common/neuter distinction and seeks to discover the way in which the phenomenon is structured, examples such as in 6 are included, and their behavior can indeed be shown to relate to other examples of the semantically-driven use of the neuter pronoun.

A final category of cases in which no agreement relation can be established includes idioms, collocations, and formulaic sequences such as *het smaakt* ‘it tastes [good]’ or (about the weather) *het ziet er goed uit* ‘it looks good’, and cases where the antecedent could not clearly be determined. These are not further discussed below.

In the remaining cases, the correct resolution of the coreferentiality relations—whether these were grammatical or semantic—is a necessary premise for obtaining reliable results. Most instances in the corpus proved unproblematic during annotation: A sample of unproblematic instances was crosschecked by a second annotator, which yielded no disagreement. In certain cases, several interpretations were indeed possible. Instances deemed to be problematic by the annotator were evaluated by a second annotator and discussed until agreement was reached.

### 3.2. Factors in the Analysis.

The entire procedure yielded a database of 1,731 tokens, which were subsequently tagged for a large number of factors. Table 2 provides an overview.

<b>Factor</b>	<b>Factor levels</b>
Lexical.gender	Feminine Masculine
Noun.frequency	Numerical value ranging from 0.69 to 7.87
N.morphology	Feminine.morphology Not.marked
N.semantics	Complex Simple
Antecedent.definiteness	Definite Indefinite Zero
Antecedent.function	Comp(lement) Object Oblique NP.part Subject Incomp(lete.sentence)
Pronoun.sentence	Cop.sentence Obj.trans.V Subj.lex.V Ambig.incompl
Pronoun.weight	Full Reduced
Syntactic.distance	Beyond.turn In.turn In.sentence
Speaker.gender	Female Male
Speaker.residence	West-Flanders Other.region (East-Flanders, Brabant, Limburg)
Speaker.birthyear	Oldest.generation (born before 1960) Youngest.generation (born in 1960 or later)
Speaker.level	High Middle Low Unknown

Factor	Factor levels
Register	SpontPrivDia SpontPubDia SpontPubMono PrepPubMono

Table 2. Overview of all fixed factors included in the statistical analysis.

Most of the factors and factor levels mentioned in table 2 were taken from the literature where they have been shown to influence the choice between lexical and semantic agreement in Dutch. As noted above, the list of factors includes discourse contextual effects, which often resist straightforward quantification. The investigation therefore focuses on factors that allow for operationalizations with levels that can be determined without analyzing longer stretches of discourse. In addition, the factors adopted here do not yield too many missing values. Most factors were already described in section 2; their concrete operationalization is explained in the remainder of this section in the order in which they appear in table 2.

There were three factors directly related to the antecedent noun. First, we used *Lexical.gender*, that is, masculine versus feminine, as described by the *Woordenboek der Nederlandsche Taal* (Dictionary of Dutch Language).<sup>6</sup> The second factor was *Noun.frequency* (implemented as a logarithmic transformation of the frequency in the Spoken Dutch Corpus).<sup>7</sup> Finally, we considered whether the noun contained a morphological element marking feminine gender, such as *-ing* in *herstelling*

<sup>6</sup> The (historical) *Woordenboek der Nederlandsche Taal* was preferred over the (contemporary) Van Dale dictionary because it provides more explicit information on variation in nouns' gender. For example, *boek* 'book' is neuter in Standard Dutch but masculine in some southern, nonstandard varieties. For nouns with variable gender, gender was determined using gender clues in the adnominal domain, such as the nonstandard indefinite masculine article *ne*. With this procedure, it was possible to determine the gender of all nouns in the sample.

<sup>7</sup> Despite the fact that fewer data are included in the Spoken Dutch Corpus, they were preferred over SUBTLEX-NL (Keuleers et al. 2010), because they are drawn from the same corpus as the pronouns, and therefore yield no missing values.



reparation' or *-de/-te* in *grootte* 'size' (*N.morphology*).<sup>8</sup> For noun semantics (*N.semantics*), a distinction between "simple" and "complex" was drawn (see above for discussion), the latter functioning as a residual category to which the antecedents were assigned based on a negative semantic criterion, that is, if they were noncount, abstract, or both. Both features inherent to the semantics of the referent and to the referential context were captured in one factor in the process.<sup>9</sup>

In addition, we encoded a number of factors related to the discourse context in which the antecedent was used. The determiner preceding the noun (*Antecedent.definiteness*: indefinite article, other indefinite determiners, definite article, demonstrative, possessive) was used to operationalize the antecedent's pragmatic prominence. Moreover, the antecedent's syntactic function was encoded (*Antecedent.function*), distinguishing between, essentially, three categories from Keenan & Comrie's (1977) Accessibility Hierarchy (that is, subjects, objects (direct and indirect), obliques), and three additional categories that occurred several times in the data, namely, antecedents that functioned as complements or as nonarguments, antecedents that served a modifying function within another constituent, and a category labeled "incomplete" that contained antecedents in fragmented or incomplete sentences, which rendered it impossible to define their grammatical function.

As for the pronoun and the clause in which it is used (*Pronoun.sentence*), the frequent occurrence of *het/'t* 'it' in copular

---

<sup>8</sup> There are also nouns that are morphologically marked for masculine gender, in particular, derivations with the *-er* suffix, which were too infrequent to be included as a separate category: The corpus only contains ten pronouns referring to six nouns (*ontstopper* 'plunger', *omvormer* 'transformer', *gsm-lader* 'cell phone charger', *pijnstiller* 'painkiller', *brander* 'burner'), and 12 additional pronouns referring to loan words (in particular, *printer*, *scanner*, *chronometer*).

<sup>9</sup> The literature on semantic gender mentions different behavior of nouns referring to collectives (Geerts 1983, Haeseryn et al. 1997, §3.3.3.6) and to substances (Van Haeringen 1936), which are associated with feminine pronouns. In the Belgian part of the Spoken Dutch Corpus, only the first phenomenon can be observed. It is rare, however, as most collectives show a general preference for the neuter pronoun, which does not significantly differ from the preference observed in other nouns. Therefore, these semantic parameters were not taken into consideration (which means that the instances were included in the category of the complex entities).

clauses has been pointed out several times (for example, Romijn 1996; see section 3.1 above for discussion). For that reason, copular clauses were included as a separate category. The pronoun's syntactic function (subject versus object) is known from the literature to be a relevant factor (Audring 2009:162–164). Additionally, semantic factors relevant for pronominal gender were suggested in historical varieties of English (for example, Grund 2011), but further classifications of verbs resulted in unacceptably large numbers of missing values in the data (see, however, below for some discussion of subjects of activity versus state verbs). Thus, a distinction was encoded between subject pronouns in copular clauses, other subjects (that is, subjects of lexical verbs), and objects (of transitive verbs). Moreover, pronouns in utterances without verbs were taken to be a residual category that also included ambiguous cases.

Furthermore, Dutch systematically distinguishes between full and reduced pronominal forms, and this distinction is relevant for the referential potential of the pronouns, since all full forms—apart from the 3rd person masculine and neuter pronouns—tend to be used to refer to persons only. Hence this factor (*Pronoun.weight*) was included as well, even though its relevance for resemanticization is not entirely clear.

As a final linguistic factor, the interval between the antecedent and the pronoun was included (*Syntactic.distance*), using a three-way distinction between elements found in the same sentence, in the same turn, or beyond the turn of the speaker.

As for the social factors, the Spoken Dutch Corpus provides for the possibility to include (an operationalization of) the traditional factors, namely, age (*Speaker.birthyear*), gender (*Speaker.gender*), and social class (in the guise of educational level, *Speaker.level*). The age factor, namely, speaker's birth year, yielded too many different levels to be suitable for our statistical model, and was recoded in terms of birth year before/after 1960. Because variation has been documented within Belgium with respect to the resemanticization process (De Vogelaer & De Sutter 2011), a geographical factor (*Speaker.residence*) was included as well, which distinguished between the major dialect regions in Belgium (West Flanders, East Flanders, Brabant, Limburg). Since the Spoken Dutch Corpus contains data gathered from a range of settings, it is also possible to investigate whether the choice between lexical and semantic agreement is subject to stylistic variation (*Register*). Fourteen components of the Spoken Dutch Corpus have been reduced to four

broader categories so as to reflect an increasing “attention-paid-to-speech” in the sense of Labov (1972), namely, prepared public monologues, spontaneous public monologues, spontaneous public dialogues, and spontaneous private dialogues.

#### **4. Multivariate Analysis.**

For the statistical analysis, the choice of a neuter versus non-neuter pronoun was considered as the response: Since our sample consisted of non-neuter antecedents referring to inanimates, this choice could be unambiguously linked to semantic or lexical agreement, respectively. This operationalization abstracts away from the masculine/feminine distinction that is still maintained by many speakers in the corpus. In fact, in contrast to northern Dutch, where feminine pronouns are no longer used in reference to inanimates, southern Dutch has preserved this distinction fairly well: Almost 90% of the instances of lexical agreement with feminine antecedents in the corpus consist of feminine pronouns (that is, 357 feminine pronouns versus 50 masculine ones; compare with 423 neuter pronouns qualifying as instances of semantic agreement). Overall, lexical agreement is found in a slight majority of our corpus examples (980/1,731, or 57%). Thus, the total proportion of semantic agreement (751/1,731, or 43%) does not differ too much from Audring’s (2009:167) figure of 53%.<sup>10</sup>

A mixed-effects generalized linear model was fitted (using RStudio 1.1.383; R Core Team 2016), consisting of the fixed factors mentioned in table 2 as well as two random factors (speaker, lexical item). In order to identify which of the factors in table 2 significantly contribute to the choice between a neuter and non-neuter pronoun, we adopted a bidirectional selection procedure, with AIC (that is, the so-called Akaike information criterion) as the decisive criterion. Additionally, the results of this analysis show the relative effects of each of the significant factors and the explanatory and predictive power of the model (that is, the

---

<sup>10</sup> This is a total score calculated from the data for objects and both specific and unspecific mass nouns from Audring’s (2009) “restricted data set.” While this figure is, in our view, the best measure in her data to compare with our general results, it provides far from a perfect match, as Audring includes slightly different cases in her study, such as neuter antecedents (as agreement controllers) and relative pronouns (as agreement targets).

combined effect of all significant factors). The results of the generalized linear mixed-effects model are shown in table 3 and figure 1. Before discussing these results, it is important to note that the statistical model is significantly more accurate than an intercept-only model ( $p < 0.0001$ ) and yields an  $R^2$  of .35 and a c-index of 0.97. This shows that the model has a high explanatory and predictive power; there is no multicollinearity observed.<sup>11</sup>

As can be seen in table 3, most of the significant factors in the model are linguistic, namely, the semantics of the antecedent noun (*N.semantics*), the presence of a suffix marking feminine gender on the noun (*N.morph*), the determiner combining with the antecedent (*antecedent.determiner*), the type of sentence in which the pronoun is used (*pronoun.sentence*), pronouns being either full or reduced (*pronoun.weight*), and the interval between antecedent and pronoun (*syntactic.distance*). For antecedent function, only the residual category of incomplete clauses triggers significantly more neuter pronouns. Of the linguistic factors, the antecedent's lexical gender (masculine or feminine) and the usage frequency of the noun did not have any effect. Additionally, two social factors turn out to be significant, namely, speech register and speaker gender. Speaker residence (namely, the traditional dialect regions West Flanders, East Flanders, Brabant, and Limburg) and age (in the form of birth year before/after 1960) turn out not to be statistically significant factors. Note, however, that the total proportion of semantic agreement (751/1,731, or 43%) indicates that from a long-term perspective, resemanticization is also taking effect in Flanders.

Table 3 shows the relative effect of each of the factors (the further away from 1, the stronger the effect) and the effect direction (does factor level  $x$  stimulate neuter or non-neuter pronouns?). Effect size and direction are measured by means of reference coding: One of the factor levels of each factor is used as reference level, against which the effect of the other factor levels is measured. The reference values are listed in parentheses in table 3. Additionally, effects are also depicted in figures 1–9.

---

<sup>11</sup> The full statistical model is provided at <http://hdl.handle.net/1854/LU-8652787>

N.semantics (complex)			
	simple	16.67	Non-neuter pronoun
N.morph (unmarked)			
	fem.morph	2.95	Neuter pronoun
antecedent.function (subject)			
	incompl.	3.17	Neuter pronoun
antecedent.determiner (demonstrative)			
	indef.art	3.27	Neuter pronoun
	zero	3.67	Neuter pronoun
pronoun.sentence (obj.trans.V)			
	subj.lex.V	3.51	Neuter pronoun
	cop.sentence	8.04	Neuter pronoun
	ambig.incompl	8.78	Neuter pronoun
pronoun.weight (reduced)			
	full	2.59	Neuter pronoun
syntactic.distance (beyond.turn)			
	in.turn	1.81	Non-neuter pronoun
	in.sentence	3.13	Non-neuter pronoun
speaker.gender (male)			
	female	1.92	Neuter pronoun
register (SpontPrivDia)			
	SpontPubDia	5.88	Non-neuter pronoun
	SpontPubMono	3.70	Non-neuter pronoun
	PrepPubMono	33.33	Non-neuter pronoun

Table 3. Overview of all significant factors, their effect sizes and directions.

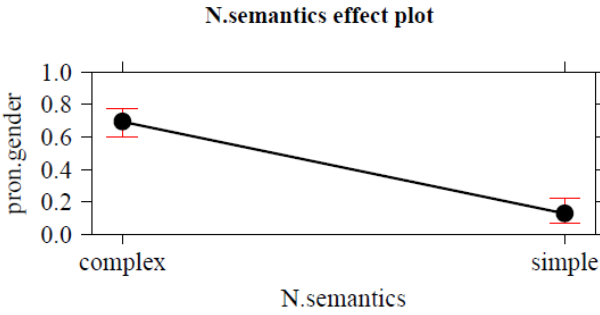


Figure 1. Effect plot: *N.semantics*.

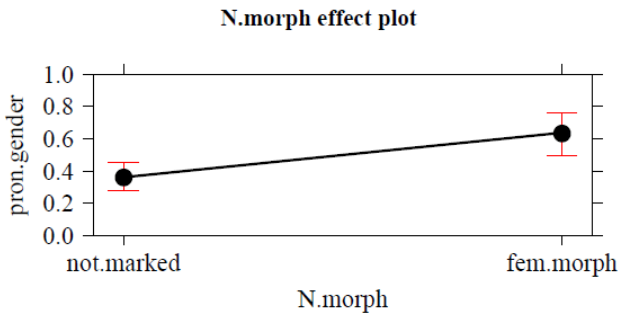


Figure 2. Effect plot: *N.morph*.

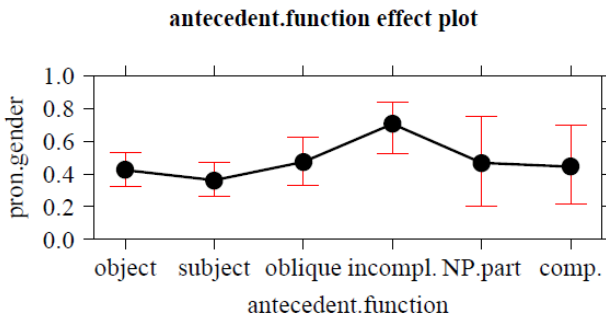
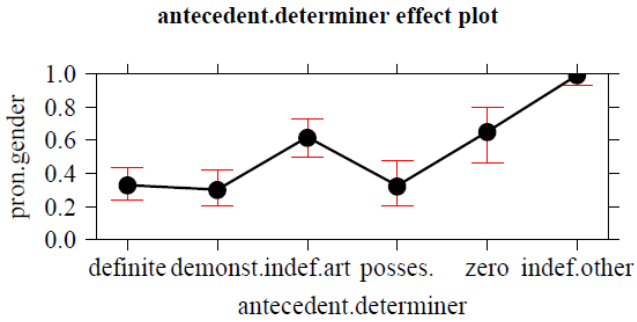
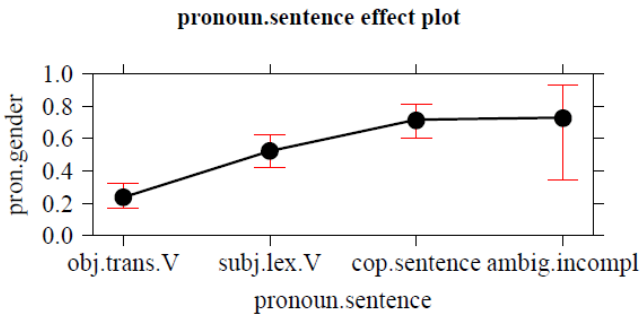
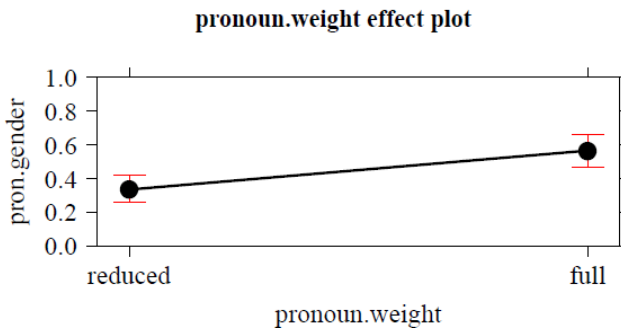


Figure 3. Effect plot: *Antecedent.function*.

Figure 4. Effect plot: *Antecedent.determiner*Figure 5. Effect plot: *Pronoun.sentence*.Figure 6. Effect plot: *Pronoun.weight*.

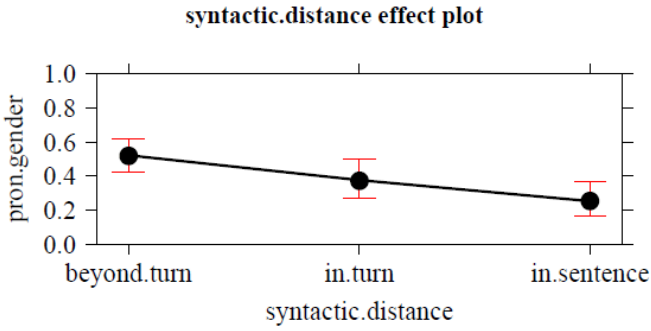


Figure 7. Effect plot: *Syntactic.distance*.

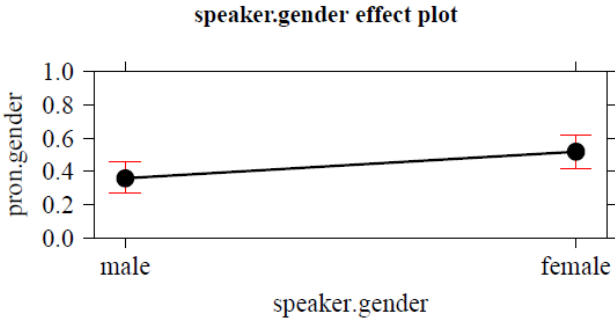


Figure 8. Effect plot: *Speaker.gender*.

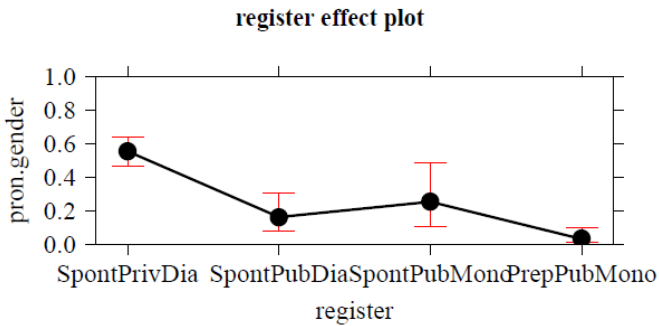


Figure 9. Effect plot: *Register*.



A factor that has been included in the corpus studies by both Audring (2009) and Bouma (2018) is noun semantics. Both studies found that lowly individuated inanimates trigger significantly more semantic agreement (neuter pronouns) than highly individuated ones. In our results, the distinction between nouns with simple and complex semantics, which is an operationalization of the role of the Individuation Hierarchy in 2, is the second most powerful factor (the odds ratio of 16.67 indicates that neuter *het*/'it' is 16.67 times more likely to be used for nouns with complex semantics). An additional factor related to the noun is morphological marking for feminine gender. The effect of being morphologically marked for feminine gender is somewhat unexpected: Morphologically marked feminine nouns trigger more semantic agreement, despite the predictability of their gender. This may be attributable to the fact that many of the suffixes involved (for example, *-heid* as in *gezondheid* 'health' or *-de/-te* as in *grootte* 'size') are used to build nouns with highly abstract meanings.

Additional significant factors not directly related to the antecedent noun but rather to the context in which it is used, are its syntactic function and the determiner with which it is combined. The former factor was introduced to test the relevance of Keenan & Comrie's (1977) Accessibility Hierarchy for the alternation between lexical and semantic agreement. However, the only significant difference is a higher propensity for semantic agreement in incomplete sentences, which is unexpected because these were seen as a residual category. As for the determiner preceding the antecedent, indefinite NPs (zero, indefinite article, other indefinite determiners) trigger more semantic agreement than definite NPs (definite article, demonstrative, possessive; see Bouma 2018 on the differences between NPs introduced by indefinite *een* 'a' and definite *het* 'the').

Another set of factors in table 3 concern the pronoun and the sentence in which it is used. As for the weight of the pronoun, recall from section 2 that no clear predictions could be formulated as to the type of agreement expressed with full vis-à-vis reduced pronouns. It turns out that full forms are more commonly used for semantic agreement in our data, which is in line with Audring's (2009:162) observations for northern Dutch. As for the sentence in which the pronoun is used, semantic agreement is least likely to appear when the pronoun serves the function of grammatical object in the sentence.

Slightly more semantic agreement is found when the pronoun serves as a subject, both of a lexical verb and, especially, of the copula *zijn* ‘be’ combined with a predicative adjective. As in the antecedent clause, the verb’s contribution to the individuation of a pronoun’s referent is expected. It relates, for instance, the degree to which a described activity entails an instigator. Since the subject of a copular clause typically ranks low with respect to agentivity, the preference of copular clauses for neuter *het* / ‘it’ is obvious. The differences between clauses with object pronouns and subject pronouns are less expected. They may occur because the category of the subject pronouns includes both highly agentive subjects, whose referents typically rank high on the Individuation Hierarchy, and subjects with lowly individuated referents, such as subjects of state verbs. We further analyzed a sample of the data (the spontaneous conversations between peers, that is, the A-component of the Spoken Dutch Corpus) distinguishing between pronouns used as subjects of activity versus state verbs. This analysis indeed yielded very different agreement preferences, with 29% (19/65) of subjects of activity verbs showing semantic agreement and 55% (109/197) with state verbs. The latter figure compares well with the proportion found for objects, which show 56% (200/355) semantic agreement. While these numbers provide a clear indication that the semantics of the predicate also plays a role in the alternation between semantic and lexical agreement, a more precise investigation of this role is left for further investigation.

Apart from factors concerning the antecedent and the pronoun by themselves, the syntactic interval between the two was expected to play a role as well. This prediction is in line with the Agreement Hierarchy (Corbett 1979), which has been found to be relevant in much work on agreement, including Audring’s (2009) study of resemanticization in the Hollandic part of the Spoken Dutch Corpus. Table 3 contains a rather rough operationalization of the distance between the pronoun and its antecedent: It distinguishes between pronouns and antecedents used in the same sentence, the same turn, and beyond turns. The data in table 3 indeed show that semantic agreement occurs more often when the pronoun and its antecedent are further apart.

When it comes to social factors, female speakers use more semantic agreement than male ones, which confirms the common observation that women, on average, tend to be ahead of men in processes of language change (an effect also found in Bouma 2018). The most important factor

in the entire analysis is a social one, too, namely, register: Semantic gender is more often found in less formal registers, that is, in spontaneous speech settings. This result also corroborates a finding from Bouma's (2018) Twitter data. Thus, semantic agreement correlates with less attention being paid to speech, which confirms the hypothesis formulated at the end of section 2, namely, that resemanticization qualifies as a change from below in Labovian terms.

All in all, then, table 3 provides further evidence for the role played in resemanticization by well-investigated factors, such as noun semantics and the interval between antecedent and pronoun, and adds factors for which less empirical evidence has been presented in the literature, including the antecedent's determiner, the predicate of the pronoun sentence, and the social factor *register*.

## 5. Discussion.

In this section, we attempt to provide a coherent account of how the factors described in section 4 contribute to the variation between lexical and semantic agreement. This is done by linking their effect on pronominal gender to aspects of linguistic cognition as described under the header LEXICAL ACCESS to be defined below (Levelt 1989 and subsequent work). In an article discussing usage data, the link between resemanticization and lexical access can only be considered tentative; hence, the hypotheses put forward in the remainder of this section are somewhat speculative in nature. However, these hypotheses provide a clear research agenda for future experimental and/or crosslinguistic work.

As a first step, some differences from the northern Dutch situation described by Audring (2009) need to be brought to the fore. Indeed, northern Dutch seems to be on its way to replace lexical agreement by a semantic system operating on the basis of individuation, in particular a distinction between mass and count nouns. At first glance, southern Dutch pronominal gender differs from the pronominal gender in northern Dutch: whereas the North shows generalized use of masculine or common pronouns for simple entities irrespective of their gender, neuter nouns referring to inanimates in the South always trigger neuter pronouns. In this respect, southern Dutch agreement more strongly resembles the historical system. A similar observation can be made with respect to the masculine/feminine distinction, which is still quite robust in the South, but more or less gone in the North. On closer inspection,

however, the extent to which semantic agreement is observed in our data does not differ substantially from Audring's (2009:167) counts for northern Dutch. Also, even if southern Dutch does not show evidence of a count/mass dichotomy, semantic gender also operates on the basis of the more general property of individuation. The main difference between northern and southern Dutch seems to be that in southern Dutch, the cut-off point between referents triggering neuter vis-à-vis common pronouns lies higher on the Individuation Hierarchy in 2, meaning that the neuter pronoun turns into the default personal pronoun for referring to inanimate entities, irrespective of the lexical gender of the nouns involved. Such a system has also been documented for medieval English (Siemund & Dolberg 2011). Theoretically speaking, then, the near-exclusive usage of *het/'t* 'it' for neuter nouns in southern Dutch can also be due to the fact that lexical and semantic agreement yield the same outcome.

Apart from differences, both northern and southern Dutch show a link between the neuter pronoun and low individuation. This may have been inherited from older stages of Dutch, via usages of the neuter pronoun other than referring to lexically neuter nouns: Indeed *het/'t* 'it' is also used to refer to non-nominal antecedents denoting activities, states of affairs, events or propositions (Haeseryn et al. 1997, §5.2.5.2.2, Kraaikamp 2012:208–213). The semantics of *het/'t* 'it' allows the pronoun to be used for all concepts that have not been identified in previous sentences, or for referents that are backgrounded in discourse. A closely parallel development has been observed in Old English, which also displayed a tendency to use the neuter pronoun to refer to specific objects as well as less specifically defined ideas mentioned in the discourse (such as a predicate or the entire previous sentence; Curzan 2003:92). In English, neuter ultimately developed into a default pronoun for virtually all inanimate references, which mirrors the southern, rather than the northern Dutch situation.

Apart from semantic affinities between the neuter pronoun and lowly individuated nouns, there is an additional factor making them especially susceptible for resemanticization. Lowly individuated nouns typically combine with a zero determiner, which is not gender-marked. This may cause their gender to be less deeply entrenched in the mind of the speaker. Section 4 also provided evidence that the change is affected by discursive factors, such as how high the referent of the pronoun ranks on hierarchies of givenness (visible through the choice of the antecedent's

determiner), or the predicate in the pronoun sentence. Following Rose (2005), these factors can be interpreted as the relative salience of an entity in discourse. It seems plausible that salience in discourse has an impact on the agreement choice, which is indicative of an underlying—most likely psycholinguistic—process. A suitable concept to help explain this process would be lexical access. Lexical access is defined as the way in which individuals retrieve words from their mental lexicon (Field 2003), namely, the process by which the sound-meaning connections of language (that is, lexical entries) are activated.

Levelt (1989) suggests that lexical entries contain two types of information that allow individuals to recognize and understand words: information about the form and about the meaning. The form component of lexical entries refers to phonological and morphological information, while the meaning component refers to the syntax and semantics (Field 2004). Lexical gender is considered to be an inherent property of the noun, stored within its lexical entry.

Research on lexical access consistently shows that the ease of accessing a lexical entry depends on its preaccess level of activation which, in turn, depends on two factors, namely, the resting level of activation and context. The resting level of activation is primarily a function of the frequency of access: The more an entry is accessed, the higher its resting level. With respect to context, the resting level depends on the other entries that are recently accessed. Many of the factors that affect lexical access are strikingly similar to the factors involved in the alternation between lexical and semantic agreement: noun semantics, lexical ambiguity, and context effects.<sup>12</sup> The production of a gender-marked pronominal form also entails the lexical selection of the referent noun (Navarrete & Costa 2009). If the masculine/feminine distinction in Dutch is becoming blurred, and if formal exponents of gender are lacking in many usage contexts, it seems likely that language learners face increasing difficulty in deriving a noun's gender from the input. This also

---

<sup>12</sup> With respect to noun semantics, see Paivio 1991 and Jessen et al. 2000 on word concreteness and imagery, which, having to do with the ability to visualize lexical items, are concepts clearly related to individuation. With respect to lexical ambiguity, see MacDonald et al. 1994 on how readers and listeners retrieve the contextually appropriate meaning of lexical items that have multiple meanings.

becomes clear in the adnominal domain, where the acquisition of target-like gender markers is much slower in Dutch than in related languages, both in first and second language acquisition (Blom et al. 2006, Cornips & Hulk 2006). Correspondingly, the process of retrieving gender information from the memory is hampered and, as evidenced by processing studies, slowed down (Loerts 2012, Brouwer et al. 2017). With regard to pronominalization in language usage, this results in more cases of semantic agreement, and means that the alternation between lexical and semantic agreement depends on factors that facilitate or inhibit gender retrieval. One of the inhibiting factors is individuation: Simple entities have a high degree of so-called concreteness or imagery, making their lexical entry more easily accessible and the gender information more easily retrievable. This is also where contextual effects such as the ones found in this study can be observed: Given that the aim of accessing a lexical entry in the memory is to select the appropriate grammatical gender in order to be able to produce the agreeing gender-marked pronominal form, the ease of activation will be greater with nouns that are already salient in discourse. These contextual effects include the distance between antecedent and pronoun, which, on average, negatively influences an antecedent's salience.

Finally, an account relying on the concept of lexical access may also explain why the results for the more formal registers differed significantly from those for the informal registers: Antecedents occurring in the most informal part of the corpus were more often referred to by semantically agreeing pronouns. A higher degree of monitoring (in the formal registers) is likely to increase overall levels of salience, which is shown to yield a higher proportion of grammatical agreement.

## **6. Conclusion.**

The aim of this article was to supplement Audring's (2009) account of the resemanticization of Dutch pronominal gender, investigating a region where the traditional three-gender system is still found. Our quantitative analysis is more comprehensive than previous ones as it includes not only most factors known from previous analyses, but also a number of discourse factors that were only hinted at in the literature. Even though a quantitative study only allows for rather rough operationalizations, some of these discourse factors proved to be significant for the choice in favor of semantic vis-à-vis lexical agreement. The effects of social factors such

as gender and origin of the speakers were also confirmed. Our approach also allowed for a more fine-grained analysis of register effects than possible with Bouma's (2018) Twitter data and found these effects to be significant.

It was speculated that the effect of many factors can be understood as a function of the ease with which a lexical entry and its gender can be accessed by speakers (see Levelt 1989). In attributing the effects to lexical access, linguistic and social factors were integrated in a psycholinguistic account in which resemanticization is seen as a change from below, or transmission (De Vogelaer 2009), caused by gender distinctions becoming increasingly invisible. Of course, this account deserves to be tested on other processes of resemanticization observed crosslinguistically (see Siemund 2002, Fernández-Ordóñez 2009 for more examples).

Specifically for Dutch, the fact that southern varieties, unlike northern varieties such as Standard Dutch, still, to some extent, distinguish between masculine and feminine gender, casts serious doubt on the received history of Dutch pronominal gender. Geerts' (1966) influential account describes 17th-century Dutch as a dyadic system with a solid distinction between common and neuter nouns, both in the adnominal and pronominal domain. This description is echoed in most reference grammars of Dutch (for example, Haeseryn et al. 1997). However, if the use of neuter pronouns for non-neuter antecedents is caused by uncertainty with respect to the masculine/feminine distinction, as claimed in this article, one would rather expect a direct transition from the three-gender system with lexical agreement in pronouns to the semantic agreement systems found in contemporary Dutch, without an intermediate stage as proposed by Geerts (1966). This is also the conclusion drawn by Kraaikamp (2017:142–143) in her study of semantic agreement in historical cookbooks.

On a synchronic level, the proposed account also yields predictions as to the processing of gendered pronouns. These concern all the factors listed in table 3, for which, in principle, semantically motivated neuter pronouns are likely to be less strongly perceived as gender mismatches in conditions favoring semantic agreement in Dutch. Further predictions concern consistent, three-gender varieties of West-Germanic (for example, Standard German), in which one would, for instance, expect semantically motivated neuter pronouns to be less strongly perceived as a gender

mismatch than other mismatches in the pronominal domain (for instance, feminine pronouns for masculine nouns and vice versa). There is some experimental evidence that this is indeed the case (De Vogelaer et al. 2020), which would then also provide an explanation for the fact that resemanticization has been able to remain under the radar for so long.

#### DATA SOURCES

*Corpus Gesproken Nederlands* [Spoken Dutch Corpus]. Available at <http://lands.let.ru.nl/cgn/>.  
*Taalunie*. Available at [taalunieversum.org](http://taalunieversum.org), accessed in January, 2018.  
*Woordenboek der Nederlandsche Taal* [Dictionary of Dutch Language].  
*Historisch Wetenschappelijk Beschrijvend Woordenboek van het Nederlands van 1500 tot 1976*. Available at <http://ivdnt.org/onderzoek-a-onderwijs/lexicologie-a-lexicografie/WNT>.

#### REFERENCES

- Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: Routledge.
- Arnold, Jennifer. 1998. *Reference form and discourse patterns*. Stanford, CA: Stanford University dissertation.
- Audring, Jenny. 2006. Pronominal gender in spoken Dutch. *Journal of Germanic Linguistics* 18. 85–116.
- Audring, Jenny. 2009. *Reinventing pronoun gender*. Utrecht: LOT.
- Audring, Jenny. 2014. Gender as a complex feature. *Language Sciences* 43. 5–17.
- Audring, Jenny, & Geert Booij. 2009. Genus als probleemcategorie. *Perspectieven op het genus in het Nederlands (Taal en Tongval, Themanummer 22)*, ed. by Leonie Cornips & Gunther De Vogelaer, 13–37. Tongeren: Michiels.
- Blom, Elma, Daniela Poliřanská, & Fred Weerman. 2006. Effects of age in the acquisition of agreement inflection. *Morphology* 16. 313–336.
- Bouma, Gosse. 2018. Agreement mismatches in Dutch relatives. *Belgian Journal of Linguistics* 31. 137–164.
- Brouwer, Susanne, Simone Sprenger, & Sharon Unsworth. 2017. Processing grammatical gender in Dutch: Evidence from eye movements. *Journal of Experimental Child Psychology* 159. 50–65.
- Corbett, Greville G. 1979. The agreement hierarchy. *Journal of Linguistics* 15. 203–224.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Cornips, Leonie, & Aafke Hulk. 2006. External and internal factors in bilingual



- and bidialectal language development: Grammatical gender of the Dutch definite determiner. *L2 acquisition and creole genesis. Dialogues*, ed. by Claire Lefebvre, Lydia White, & Christine Jourdan, 355–378. Amsterdam: John Benjamins.
- Curzan, Anne. 2003. *Gender shifts in the history of English*. Cambridge: Cambridge University Press.
- De Schutter, Georges, Boudewijn van den Berg, Ton Goeman, & Thera de Jong. 2005. *Morphological atlas of the Dutch dialects*, vol. 1. Amsterdam: Amsterdam University Press.
- De Vogelaer, Gunther. 2009. Changing pronominal gender in Dutch: transmission or diffusion? *Language variation, European perspectives II*, ed. by Stavroula Tsiplakou, Marilena Karyolemou, & Pavlos Pavlou, 71–80. Amsterdam: John Benjamins.
- De Vogelaer, Gunther, & Gert De Sutter. 2011. The geography of gender change. Pronominal and adnominal gender in Flemish dialects of Dutch. *Language Sciences* 33. 192–205.
- De Vogelaer, Gunther, Sarah Schimke, Greg Poarch, Johanna Fanta, & Lukas Urbanek. 2020. Syntactic or semantic gender agreement in Dutch, German and German learner Dutch: A speeded grammaticality judgement task. *German and Dutch in contrast: Synchronic, diachronic and psycholinguistic perspectives*, ed. by Gunther De Vogelaer, Dietha Koster, & Torsten Leuschner, 271–298. Berlin: de Gruyter.
- De Vos, Lien, & Gunther De Vogelaer. 2011. Dutch gender and the locus of morphological regularization. *Folia Linguistica* 45. 245–281.
- Durieux, Gert, Walter Daelemans, & Steven Gillis. 1999. On the arbitrariness of lexical categories. *Computational linguistics in the Netherlands 1998*, ed. by Frank Van Eynde, 19–35. Amsterdam: Rodopi.
- Eerten, Laura van. 2007. Over het Corpus Gesproken Nederlands. *Nederlandse Taalkunde* 12. 194–215.
- Fernández-Ordóñez, Ines. 2009. The development of mass/count distinctions in Indo-European languages. *Grammatical changes in Indo-European languages*, ed. by Vit Bubenik, John Hewson, & Sarah Rose, 55–68. Amsterdam: John Benjamins.
- Field, John. 2003. *Psycholinguistics: A resource book for students*. London: Routledge.
- Field, John. 2004. *Psycholinguistics: The key concepts*. London: Routledge.
- Fletcher, William H. 1987. Semantic factors in Dutch gender choice. *Papers from the Second Interdisciplinary Conference on Netherlandic Studies*, ed. by William H. Fletcher, 51–63. Lanham, NY: University Press of America.
- Geerts, Guido. 1966. *Genus en geslacht in de Gouden Eeuw: Een bijdrage tot de studie van de nominale klassifikatie en daarmee samenhangende adnominale flexievormen en pronominale verschijnselen in Hollands taalgebruik van de*

- zeventiende eeuw. Brussel: Belgisch Interuniversitair Centrum voor Neerlandistiek.
- Geerts, Guido. 1983. Is het gemeentebestuur haar boekje te buiten gegaan? *Taalbeheersing in de administratie* 22. 877–890.
- Givón, Talmy. 1983. *Topic continuity in discourse: A quantitative cross-language study (Typological Studies in Language 3)*. Amsterdam: John Benjamins.
- Gordon, Peter, & Randall Hendrick. 1998. The representation and processing of coreference in discourse. *Cognitive Science* 22. 389–424.
- Gundel, Jeanette, Nancy Hedberg, & Ron Zacharski. 1993. Cognitive status and the form of referring expressions. *Language* 69. 274–307.
- Grund, Peter. 2011. The science of pronominal usage: *He* and *it* in coreference to inanimate entities in Late Middle English texts on alchemy. *Journal of English Linguistics* 39. 335–358.
- Haeringen, Coenraad B. van 1936. *Geslacht, verbuiging, voornaamwoordelijke aanduiding: Rapport uitgebracht aan Zijne Excellentie den Minister van Onderwijs, Kunsten en Wetenschappen*. 's Gravenhage: Van Goor Zonen.
- Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jaap de Rooij, & Maarten van den Toorn. 1997. *Algemene Nederlandse Spraakkunst*. Groningen/Deurne: Martinus Nijhoff/Wolters Plantyn.
- Hoppenbrouwers, Cor. 1983. Het genus in een Brabants regiolect. *TaBu* 13. 1–25.
- Horst, Joop van der 2008. *Geschiedenis van de Nederlandse syntaxis*. Leuven: Universitaire Pers.
- Jessen, Frank, Reinhard Heun, Michael Erb, Dirk Oliver Granath, Uwe Klose, Andreas Papassotiropoulos, & Wolfgang Grodd. 2000. The concreteness effect: Evidence for dual coding and context availability. *Brain and Language* 74. 103–112.
- Josefsson, Gunlog. 2006. Semantic and grammatical genders in Swedish— independent but interacting dimensions. *Lingua* 116. 1346–1368.
- Keenan, Edward L., & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8. 63–99.
- Keuleers, Emmanuel, Marc Brysbaert, & Boris New. 2010. SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods* 42. 643–650.
- Köpcke, Klaus-Michael, & David Zubin. 1983. Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache. *Zeitschrift für Germanistische Linguistik* 11. 166–182.
- 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*. Utrecht: LOT.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia, PA: 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.
- Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Loerts, Hanneke. 2012. *Uncommon gender. Eyes and brains, native and second language learners, and grammatical gender*. Groningen, The Netherlands: University of Groningen Dissertation.
- Luraghi, Silvia. 2011. The origins of the Proto-Indo-European gender system: Typological considerations. *Folia Linguistica* 45. 435–464.
- MacDonald, Maryellen C., Neal J. Pearlmutter, & Mark S. Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101. 676–703.
- Mills, Anne E. 1986. *The acquisition of gender. A study of English and German*. Berlin: Springer.
- Navarrete, Eduardo, & Albert Costa. 2009. The naming of gender-marked pronouns supports interactivity in models of lexical access. *Psicológica* 30. 301–321.
- Oostdijk, Nelleke. 2000. The Spoken Dutch Corpus. Overview and first evaluation. *Proceedings of LREC-2000*, ed. by Maria Gravididou, George Carayannis, Stella Markantonatou, Stelios Piperidis, & Gregory Stainhaouer, 887–894. Athens: International Conference on Language Resources and Evaluation.
- Paivio, Allan. 1991. Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology* 45. 255–287.
- Plevoets, Koen, Dirk Spielman, & Dirk Geeraerts. 2009. De verspreiding van de -e(n)-uitgang in attributieve positie. *Perspectieven op het genus in het Nederlands (Taal en Tongval, Themanummer 22)*, ed. by Leonie Cornips & Gunther De Vogelaer, 112–143. Tongeren: Michiels.
- R Core Team. 2016. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at <https://www.R-project.org/>.
- Romijn, Kirsten. 1996. *Hoe doen we het? Verwijzen naar linguïstische en cognitieve representaties met het voornaamwoord 'het'*. Amsterdam: Publicaties van het P. J. Meertens-Instituut.
- Rose, Ralph. 2005. *The relative contribution of syntactic and semantic prominence to the salience of discourse entities*. Chicago, IL: Northwestern University dissertation.
- Sasse, Hans-Jürgen. 1993. Syntactic categories and subcategories. *Syntax. Ein internationales Handbuch zeitgenössischer Forschung*, ed. by Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, & Theo Vennemann, 646–686. Berlin: de Gruyter.

- Semplicini, Chiara. 2012. Dutch double gender nouns: Arbitrary or motivated agreement? *Journal of Germanic Linguistics* 24. 133–186.
- Siemund, Peter. 2002. Mass versus count: Pronominal gender in regional varieties of Germanic languages. *Sprachtypologie und Universalienforschung (STUF)* 55. 213–233.
- Siemund, Peter. 2008. *Pronominal gender in English: A study of English varieties from a cross-linguistic perspective*. London: Routledge.
- Siemund, Peter, & Florian Dolberg. 2011. From lexical to referential gender: An analysis of gender change in medieval English based on two historical documents. *Folia Linguistica* 45. 489–534.
- Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. *Grammatical categories in Australian languages*, ed. by Robert Dixon, 112–171. Canberra: Australian Institute of Aboriginal Studies.
- Verheijen, Lieke. 2018. *Is Textese a threat to traditional literacy? Dutch youths' language use in written computer-mediated communication and relations with their school writing*. Utrecht: LOT.
- Verhoeven, Peter R. F. 1990. Voornaamwoordelijke aanduiding in het hedendaagse Nederlands. *De nieuwe Taalgids* 83. 494–513.
- Ward, Gregory. 1985. *The semantics and pragmatics of preposing*. Philadelphia, PA: University of Pennsylvania dissertation.
- Wurzel, Wolfgang U. 1986. Die wiederholte Klassifikation von Substantiven. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikation* 39. 76–96.

Lien De Vos

Université de Liège  
Bât. A2 Langue et linguistique néerlandaises  
Place Cockerill 3-5  
4000 Liège, Belgium

Gert De Sutter

Ghent University  
Faculty of Arts and Philosophy  
Department of Translation, Interpreting and Communication  
Groot-Brittanniëlaan 45, geb. B  
9000 Gent, Belgium  
[gert.desutter@ugent.be]

Gunther De Vogelaer  
University of Münster  
Institut für Niederländische Philologie  
Alter Steinweg 6/7  
48143 Münster  
Germany  
[devogelaer@uni-muenster.de]