Let's Agree to Disagree.
(Variation in) the Assignment of Gender
to Nominal Anglicisms in Dutch

Karlien Franco, Eline Zenner, Dirk Speelman

QLVL, KU Leuven

In this paper, we investigate gender assignment to recently borrowed English loanwords in Dutch, introducing several innovations to the field of gender assignment to anglicisms. For example, we use multiple mixed-effects logistic regression to determine which factors underlie gender assignment in Dutch. This model indicates that there is variation in the degree of homogeneity in the speech community (that is, agreement among respondents) concerning the gender assigned to an anglicism; therefore, we analyze the contexts in which homogeneity is the lowest. Our analysis reveals that the degree of consensus does not solely depend on how established an anglicism is. In contrast to what has been argued in previous studies, gender assignment to anglicisms in Dutch is not a categorical process: Gender variation increases when respondents are faced with a conflict between the default article *de* and some factor that favors its neuter counterpart, *het*.[*]

1. Introduction.

This paper focuses on the gender assigned to English nouns when they are borrowed into a gendered language. More specifically, we analyze the factors that govern gender assignment to anglicisms in Dutch. Our analysis indicates that there is variation when gender is assigned to these nouns: Language users do not necessarily agree about the gender they assign to a borrowed noun. Furthermore, we discuss the conditions under which gender variation is the greatest.

This paper is structured in the following way. Section 2 provides an overview of previous studies on gender assignment to anglicisms. First, some universalities that have been identified in previous research are outlined. Then, the notion of gender variation is introduced. In section 3, the gender system of Dutch is discussed. The gender system that applies to native nouns is described, followed by an overview of previous research into gender assignment to anglicisms in Dutch. Section 4 discusses the questionnaire that was used for the analysis, the factors that were taken into account, and the methodology of the study. Section 5 inquires into the factors that influence the choice of the common or neuter article for an anglicism in Dutch. Furthermore, the amount of agreement among the respondents regarding this choice is analyzed. Section 6 outlines the conclusions drawn on the basis of this analysis, the shortcomings of the present study, and suggestions for further research. Section 7 is a conclusion.

2. Gender Assignment to Loanwords.

2.1. *Universalities in Gender Assignment to Loanwords*.

Gender assignment to non-native nouns has been studied in the context of lexical borrowing and against the background of bilingualism. Most studies on this topic are based on corpus data; they generally start with a set of anglicisms and their gender from a corpus, questionnaire, or dictionary (for example, Poplack et al. 1982, Budzhak-Jones 1997, Smead 2000, Violin-Wigent 2006, Cruz Cabanillas et al. 2007, Chirsheva 2009, Thornton 2009). Most of these studies rely on raw frequencies or proportions to determine the relative importance of the factors that influence gender assignment.

In general, it is assumed that the gender system of the host language applies to loanwords as well, possibly augmented by some loanword-specific additional rules (Corbett 1991, 2014). As described in Onysko 2007:164, the English noun *family*, for instance, is feminine in German, because of a native semantic rule according to which collectives of individuals receive feminine gender (for example, *die Gruppe* 'group').

Determinants of gender assignment that apply to loanwords but not to native nouns have been identified as well. Some scholars have used semantic analogy as an explanation for gender assignment (for example, Poplack et al. 1982, Corbett 1991, Thornton 2009): Anglicisms can receive a specific gender due to a semantic association with a native

noun in the host language (this factor is also referred to as "lexical equivalence", "lexical analogy" or "lexical-semantic equivalence"). More specifically, a certain gender is assigned to a loanword because a cognate or translation equivalent in the host language has that gender as well. Cruz Cabanillas et al. (2007), who use data from a corpus of English computer terms borrowed into Spanish, suggest that *URL*, for example, is feminine likely due to its association with the native feminine noun *dirección* 'address', even though most anglicisms receive masculine gender in Spanish.

However, the explanation of gender assignment through the process of semantic analogy has received some criticism (Corbett 1991, Berteloot & Van der Sijs 2003, Onysko 2007, Onysko et al. 2013). First, some scholars note that gender equivalence between semantically analogous nouns can also be explained by an underlying semantic gender association that is applicable to native nouns in the host language as well (Corbett 1991, Onysko 2007). Second, determining the closest native equivalent is often problematic. Onysko (2007:166–167, 327–328), for instance, finds that only 17 out of 63 monosyllabic masculine anglicisms have the same gender as their native German equivalents (for example, *der Beat* like *der Schlag* or *der Takt*), while for 18 out of 63 anglicisms the gender of the native equivalent is not masculine (for example, *der Song* versus *das Lied*). The rest of the 63 monosyllabic masculine anglicisms in his corpus (N=28) either have no clear German equivalent (for example, *der Flow*) or have more than one native equivalent, with different genders (for example, *der Chip* like *der Jeton*, but also *die Marke*).

Onysko et al. (2013:108–109) argue for an interpretation of lexical equivalence as a continuum of associative strength. This continuum can be interpreted in a quantitative way: The stronger the association between an anglicism and a native noun, the more likely it is that the anglicism will be assigned the gender of the associated native noun. The association between an anglicism and a native noun is the strongest when the native language has a cognate that resembles the anglicism in (etymological) form and meaning (for example, *das Notebook* analogous to German *das Buch* 'the book'). The association between an anglicism and a native noun is the weakest when the native language has a translational equivalent that is not formally related to the English noun (for example, *die E-Mail* analogous to *die elektronische Post*). The

central portion of the continuum is taken up by anglicisms associated with a basic concept for which only one native lexical item exists (for example, *die Time* like *die Zeit*).

Loanwords can also receive gender by being assigned the unmarked, or default gender of the host language (for example, Haugen 1969:440–449, Weinreich 1968:45, Hock & Joseph 1996:266–269). The unmarked gender is often equated with the most frequent gender, but, according to Corbett 1991, this view is problematic: Attributing default status to the most frequent gender can mask aspects of the underlying gender system of the host language. In Russian, for instance, the proportion of masculine nouns is exceptionally high among German loanwords (78.5% of all German loanwords are masculine). Based on frequency alone, masculine gender seems to be the default gender in Russian: The proportion of neuter native nouns is decreasing over time in favor of masculine (and feminine) nouns. So it could be argued that German loanwords receive masculine gender by default. However, Corbett argues that in this case, an explanation of gender assignment in terms of unmarked gender ignores the constraints of the host language (see also Poplack et al. 1982). The large proportion of German loans that receive masculine gender can be explained by the fact that in German, many nouns end in a consonant. These types of nouns always receive masculine gender in Russian. Finally, some scholars argue that the grammatical gender of a noun in the donor language may play a role as well (Corbett 1991; Rothe 2014). However, since in English, nouns do not carry overt gender markers, this rule does not apply to anglicisms.[1]

This study focuses on the assignment of gender to anglicisms in a gendered language, Dutch. In the Dutch language, the gender of nouns is to a large extent arbitrary. As a result, the influence of the second set of factors, which apply to loanwords but not to native nouns, may be especially significant.

2.2. Variation in the Degree of Homogeneity in the Speech Community.
The assumption that the gender system of the host language applies to loanwords, as well as the two explanations that apply to loanwords but

---

[1] Some 3rd person pronouns and wh-pronouns show gender distinctions (Quirk et al. 1985, Biber et al. 1999).

not to native nouns (that is, semantic analogy and unmarked gender) imply that speakers generally agree on the gender of loanwords (see, for instance, Corbett 1991). Any variation in gender assignment has usually been explained by the fact that the borrowed nouns are not yet established in the host language. For instance, Poplack et al. (1982) find some variation in their corpora used to study the gender of anglicisms in Puerto Rican Spanish and Montreal French. However, they explain this variation by referring to the role of the speech community: "[o]nce a borrowed noun is assigned a gender by *whatever* criteria, there is generally unanimous agreement among speakers" (Poplack et al. 1982:25).

Like the studies mentioned in section 2.1, Poplack et al. 1982 is based on corpus data; experiment-based studies on this topic are relatively few. One study that uses experimental data is Callies et al. 2012. The authors explicitly address variability in the gender of anglicisms in German. Their analysis shows that the amount of variation is relatively large, especially in experimental data. These findings suggest that in corpus-based studies, the amount of variability in the gender of loanwords could have been underestimated.

Our study aims to investigate variability in gender assignment using the following approach. First, following Callies et al. 2012, the dataset used for the analysis is based on experimental data, collected on the basis of a forced choice task, to avoid diminishing the amount of variability in the speech community. Second, we empirically investigate the findings of Poplack et al. 1982 that the gender of a loanword is variable until the noun is established in the host language by including two types of anglicisms in our dataset. On the one hand, we take into account established anglicisms, which are "widespread, recurrent and accepted" (Rothe 2014:209; see also Muysken 2000). More specifically, we use English nouns that were borrowed in the 1950s or later, but that are already listed (with a particular gender) in a Dutch dictionary. On the other hand, we incorporate nonestablished anglicisms in our dataset. These are English loans that are not dictionary-listed and not widespread. We also make sure to include nonestablished anglicisms that occur with varying frequencies. This strategy allows us to empirically assess whether there really is "unanimous agreement among speakers" (Poplack et al. 1982:25) once an anglicism becomes associated with a particular gender in the speech community.

Third, we analyze gender assignment to anglicisms in Dutch in a quantitative way. Our goal is to find out whether there are statistical differences between established and nonestablished anglicisms. Moreover, we aim to add some methodological innovations to this field of study by relying on inferential statistical techniques rather than on raw frequencies or proportions. Additionally, we use the predictions of our quantitative analysis to empirically investigate the connection between the establishment of an anglicism and the lack of homogeneity in the speech community. Before presenting our analysis, we first discuss the gender system of Dutch and some previous research on the gender of anglicisms in Dutch.

3. Gender in Dutch.
3.1. *The Gender of Native Nouns in Dutch*.
The gender of nouns in Dutch is largely arbitrary (Haeseryn et al. 1997). For most nouns in Standard Dutch, gender is not distinguishable on the basis of properties of the nouns themselves. Only for a small group of nouns, some formal or semantic tendencies have been described.[2] The definite article identifies the gender of a noun: Nouns preceded by *de* are common (for example, *de stoel* 'the chair'), while *het* is used for neuter nouns (for example, *het huis* 'the house').

---

[2] The gender of a small group of nouns can be determined on the basis of formal properties of the noun itself (Haeseryn et al. 1997). Nouns ending in suffixes such as *-de*, *-te*, *-heid*, *-ij*, *-ing*, *-nis*, and *-st*, for instance, are usually common (that is, combined with *de*), while nouns ending in the diminutive suffix (*-je* and its formal variants *-kje*, *-pje*, *-tje* and *-etje*) are always neuter (combined with *het*). The gender of a small number of nouns can be predicted on the basis of semantic properties (Haeseryn et al. 1997). Most names of people, animals, flowers, trees, fruits, seasons, numbers, sounds, letters, musical notes, and some musical instruments are assigned common gender. Names for the cardinal points, metals, languages, sports and games, and cities and countries are assigned neuter gender.

		common	neuter
definite article		*de*	*het*
attributive adj. in indefinite NP		*-e*	∅
demonstrative pron.	proximal	*deze*	*dit*
	distal	*die*	*dat*
relative pron.		*die*	*dat*
3rd person personal pron.	nominative	*hij* (masc.)/*zij* (fem.)	*het*
	oblique	*hem* (masc.)/*haar* (fem.)	*het*
possessive pron.		*zijn* (masc.)/*haar* (fem.)	*zijn*

Table 1. Gender in Dutch in the singular noun phrase.

Other constituents within the noun phrase—such as the attributive adjective, the 3rd person personal pronoun, and the demonstrative pronoun—can mirror the gender of a noun as well (see table 1).[3] Interestingly, the personal and possessive pronouns still reflect the traditional three-way division into masculine, feminine, and neuter gender. However, due to the loss of formal gender marking on the noun, the pronominal gender system shows variation (De Vogelaer & De Sutter 2011, Kraaikamp 2012; also see Audring 2009).

The Dutch language offers an interesting perspective on gender assignment to anglicisms. Most scholars assume that the native gender system is reflected in gender assignment to loanwords (see section 2.1). As the Dutch gender system is currently undergoing change, and gender assignment to native Dutch nouns is arbitrary to a large extent, it is expected that non-native factors, such as semantic analogy and default gender, will turn out to be important.

3.2. *The Gender of Anglicisms in Dutch.*
Previous research on the gender of anglicisms in Dutch has identified the most important tendency: Anglicisms are assigned common gender by default unless there is a reason for using neuter (Schenck 1985,

---

[3] Gender is not marked on the plural forms of articles, attributive adjectives, or pronouns.

Koenen & Smits 1992, Geerts 1996, Posthumus 1996, Verhoeven & Jansen 1996, Haeseryn et al. 1997, Berteloot & Van der Sijs 2003, Hamans 2009). Such reasons can be semantic or morphological, or they can involve some kind of analogy between the anglicism and a native Dutch noun.

First, some anglicisms are assigned common or neuter gender because they fit into a particular semantic category. Breeds of dogs, for instance, such as *bulldog* or *husky*, and drinks such as *gin*, *tonic*, or *whiskey*, are assigned common gender, while collectives such as *panel* or *team*, and sports such as *rugby* or *hockey*, are neuter in Dutch.[4] Some nouns that denote substances, such as *plastic* or *velvet*, are neuter, while others can be either neuter or common (for example, *de/het nylon*, *de/het rubber*).

Second, the morphology of the anglicism can also play a role. More specifically, suffixes can influence the gender of a loanword, especially when the foreign suffix resembles a native suffix. For instance, nouns ending in *-ing*, such as *dancing*, generally receive common gender (for instance, Dutch *de mededeling* 'announcement'), while nouns with the suffix *-ment*, such as *management*, are usually neuter (for instance, Dutch *het argument* 'argument', *het document* 'document').

Third, it has been noticed that many anglicisms in Dutch copy the gender of a closely related equivalent (semantic analogy). In some cases, the equivalent is a cognate, which is both formally and semantically related. For instance, *arthouse* is probably assigned neuter gender under the influence of its neuter cognate *huis* 'house'; *copyright* is probably neuter under the influence of neuter gender *recht* 'justice, law'. The gender of yet another group of anglicisms is said to be influenced by the gender of a translational equivalent, which does not formally resemble the borrowed noun. For example, it is possible that *approach* is assigned common gender by analogy with the Dutch common gender noun *aanpak*, whereas *bacon* may be neuter because it is associated with the native Dutch neuter noun *spek*.

However, it is not very clear how these rules interact. Most scholars merely provide an overview of all the rules applicable to anglicisms in

---

[4] Names of sports, such as *het voetbal* 'soccer', which are generally neuter in Dutch, occasionally occur with a common gender definite article in colloquial language.

Dutch, based on a list of English loans that follow one (or more) of those rules. Geerts (1996) is one notable exception: He constructs a hierarchy that describes what happens if multiple rules apply to a single anglicism in Dutch. According to Geerts, the gender of a cognate has a stronger influence than the prototypical gender of nouns within the same semantic field; the latter, in turn, has a stronger influence than the gender associated with a particular suffix, which is, again, more important than a translation equivalent's gender. For instance, *gingerbeer* is a neuter noun in Dutch because its association with the native neuter cognate *bier* takes precedence over the prototypical common gender associated with the semantic field *drinks*.[5]

4. Data and Methods.
Our analysis aims to answer two distinct research questions. First, we use inferential statistical techniques to determine which factors had a significant influence on the gender of anglicisms in Dutch. Second, we used the results of our statistical analysis to identify anglicisms whose gender causes the most disagreement within the speech community (that is, among the respondents of our survey). To answer these research questions, we used data collected through a questionnaire. The analysis was carried out using R (R Development Core Team 2013). This section presents the data, variables, and methodology used in the analysis. Section 4.1 outlines the design of the questionnaire. Section 4.2 provides an overview of the variables that were used for the analysis. In section 4.3, the methodology is explained.

4.1. Data Collection.
The questionnaire consisted of two parts. The first part contained a forced choice task, in which 175 sentences were presented to the respondents, with one nominal anglicism per sentence (see figure 1). Each of the 175 anglicisms in the study represents a combination of several predictors as discussed in section 4.2 below. Participants were asked to choose the definite article (common *de* or neuter *het*) and personal pronoun (masculine *hij*, feminine *ze*, or neuter *het*) to replace

---

[5] *Bier* is a cognate of the head of the compound (that is, *beer*). In Dutch, the head of a compound is usually the rightmost element of the word.

the anglicism (marked in bold) in the sentence.[6] They could also check a box labeled *Ik ken dit woord niet* 'I don't know this word' in case they did not know the anglicism in question.

(De / Het) **facelift** voor de oude dreef gaat deze week van start. (Hij / Ze / Het) moet ervoor zorgen dat het aangenamer vertoeven is in en rond het stadsbos.

☐ Ik ken dit woord niet.

'The old avenue will be getting a **facelift** starting this week. **It** will make the area in and around the town park more pleasant.'

Figure 1. Excerpt from the main section (part 1) of the questionnaire.

We collected the nominal anglicisms used in the questionnaire in two ways. The established loanwords were selected from the 14th edition of Van Dale, an important descriptive and implicitly normative dictionary of Dutch (Den Boon & Geeraerts 2008). We selected nouns with an English etymology borrowed in 1950 or later. Our final dataset contains 97 established anglicisms. The second group consists of nonestablished English nouns selected from both Dutch and English sources. As Dutch has two national varieties—Netherlandic Dutch and Belgian Dutch—and is considered a pluricentric language (Clyne 1992), we used two news corpora, the *Twente News Corpus* (TwNC) and the *Leuven News Corpus* (LeNC), which represent Netherlandic Dutch and Belgian Dutch, respectively. TwNC contains all the newspaper articles that appeared in the five national daily newspapers in the Netherlands from 1999 to 2002. LeNC is a corpus of national daily newspapers from Flanders. It contains all the newspaper data from 1999 to 2005. Together TwNC and LeNC contain over 1.6 billion words.

---

[6] The analysis of the personal pronouns is discussed in a separate contribution by the authors. In 67.14% of the cases, there is grammatical agreement between the gender of the personal pronoun and the gender of the definite article. Preliminary analyses indicate that two factors have the largest effect on the pronominal gender of the anglicism: the definite article chosen for the anglicism and the animacy of the noun.

The nonestablished anglicisms were collected by first matching all the lexemes from English WordNet to token frequency lists of both corpora (see Zenner et al. 2012, Zenner 2013). Ambiguous items, such as cognates (for example, *man*), items with unclear etymology (for example, *supervisor*), and loans from other languages (for example, *minister*) were removed from this list. Next, a manual search was conducted in the Van Dale dictionary for entries with an etymological link to English to ensure that the anglicisms were not established in Dutch. Finally, we restricted our attention to relatively frequent nonestablished anglicisms: Only anglicisms that occur with a minimum of 100 tokens in TwNC and LeNC combined and that are not listed in the 14th edition of Van Dale were included in the questionnaire. In total, our questionnaire contains 78 nonestablished anglicisms.

For drafting the stimulus sentences, we used a search engine (*google.be*) and a news website (*mediargus.be*).[7] Using these sources ensures that the sentences in the questionnaire are as natural as possible. We presented the sentences to our respondents in two randomized orders. A chi-square test indicates that there is no significant difference between the answers of the respondents in the two versions of the questionnaire.

In the second part of the questionnaire, the respondents provided information about their age, gender, place of residence, area of study, and the occupation of their parents. The data from 45 respondents were analyzed.[8] All of the respondents were students at the University of Leuven at the time the questionnaire was distributed (April/May 2013). The mean age of the respondents was 21.49. Men are underrepresented in our data (10 male respondents out of 45). The respondents come from all over Flanders: Twenty-four respondents live in the provinces of Antwerp and Flemish Brabant (central region of Flanders), 13 in the province of Limburg (in the eastern region of Flanders), and 8 in the provinces of East

---

[7] *mediargus.be* is an up-to-date collection of all Flemish newspapers and various critical Flemish magazines. It has been replaced by *gopress.be* since the distribution of the questionnaire.

[8] Forty-six students answered the questionnaire, but we excluded one respondent from the analysis because the person in question was much older than the other participants in our study (67 years old, while the other participants were between 19 and 25).

or West Flanders (in the western region of Flanders). The socio-economic status of the participants was assessed based on their parents' occupation. More specifically, we relied on the Standard Occupations Classification (*Standaard Beroepenclassificatie*) provided by the Dutch Central Bureau of Statistics in 2010 (*Centraal Bureau voor de Statistiek*). Each respondent was assigned an ID number from 1 to 45.

4.2. *Variables*.[9]

In this section, we present factors, which, based on previous research, may influence gender assignment to loanwords in Dutch. These factors are used as predictors in a logistic regression analysis. A few theoretically interesting predictors, such as the presence of a suffix associated with a particular gender, were not included in the analysis. Determining the influence of the gender associated with a suffix is relatively difficult in a multifactorial setting because most Dutch suffixes used with English loanwords are associated with common gender (one exception is the suffix -*ment*).

The first predictor in our analysis is *establishment*. As discussed in sections 2.1 and 4.1, in our study we used established and nonestablished anglicisms. Established anglicisms were then further divided into two groups, based on the Dutch article with which they appear in the dictionary: *de* versus *het*. Nonestablished anglicisms were further divided into frequent and infrequent. The values of the predictor *establishment* appear in table 2.

---

[9] The questionnaire was constructed in such a way that the degree of individuation (animate, count/mass noun, concrete/abstract noun), a factor that is especially important for pronominal gender assignment, could easily be tested, because we assumed that this factor would influence the adnominal gender of loanwords as well. However, our initial analysis showed that the individuation factor was not as important for the choice of an article; we discovered that other variables—namely, those discussed in this section—had a greater effect. Since some of these variables were only added to the dataset after the questionnaire was circulated, the distribution of the data for these variables has not been completely controlled for in the design of the questionnaire.

ANGLICISM			
established		nonestablished	
de	*het*	frequent	infrequent

Table 2. Levels of the predictor *establishment*.

In section 4.1, the different data collection methods used to collect established and nonestablished anglicisms were discussed. The frequency value of the nonestablished anglicisms is based on their token counts in TwNC and LeNC: Frequent nonestablished anglicisms occur 300 times or more in TwNC and LeNC combined, whereas infrequent nonestablished anglicisms occur less than 300 times (see appendix 2).[10] With regard to the established nouns, we follow the information listed in the Van Dale dictionary: Anglicisms listed as common nouns are coded as *established de*, while anglicisms listed as neuter nouns are classified as *established het* (see appendix 1).

The frequency of nonestablished anglicisms and the gender listed for established anglicisms are combined in the composite variable *establishment*. This categorical variable has four possible levels: "established *de*", "established *het*", "nonestablished frequent", "nonestablished infrequent".

Table 3 provides an overview of the frequency of the anglicisms in the questionnaire—that is, the number of sentences per type (or level)—along with examples. Appendices 1 and 2 contain a list of all the nouns in the questionnaire. For the established anglicisms, the gender listed in the Van Dale dictionary is specified. For the nonestablished ones, their token frequency in TwNC and LeNC, and the categorical division into frequent and infrequent is provided.

---

[10] An anonymous reviewer points out that a larger margin between the categories of frequent and infrequent anglicisms could provide more insight into the frequency effect: Nouns that occur only slightly less than 300 times are categorized differently from nouns that occur only slightly more than 300 times. However, rerunning the analysis without the nouns with a token frequency close to 300 (namely, nouns that occur between 200 and 400 times) does not cause any major differences in the results: The same variables reach significance, and the coefficients of the mixed model are almost identical to the coefficients of the model that does include these anglicisms (see section 5.1).

With respect to the established anglicisms, we expected our respondents to prefer the gender listed in the dictionary; with respect to the nonestablished anglicisms, we expected them to prefer the default article *de*. Furthermore, the predictor *establishment* allows us to determine to what extent establishment of an anglicism can explain variability in agreement about its gender (see section 2.2). We expected to observe less homogeneity in case of anglicisms that are not yet established and infrequent.

	levels	frequency in questionnaire	examples
established anglicism	established *de*	66	*blooper*, *loafer*
	established *het*	12	*aquaplaning*,[11] *jack*
nonestablished anglicism	nonestablished frequent	49	*award*, *network*
	nonestablished infrequent	48	*publicity*, *shareholder*

Table 3. Absolute frequencies of anglicisms
for the predictor *establishment*.

Our second predictor is *article of lexical-semantic equivalent*; it has two levels: *de* and *het*. This variable takes two types of analogy into account: analogy with the gender of a Dutch cognate and analogy with the gender of a Dutch translational equivalent. We define a cognate as a formally and semantically associated word: Dutch *kanaal*, for instance, is a cognate lexeme of English *channel*; Dutch *karakter* is a cognate of English *character*. A translational equivalent has a looser connection with the anglicism: The Dutch noun does not bear any formal similarity

---

[11] Preliminary analyses indicated that established anglicisms listed with *het* and established anglicisms listed with both *de* and *het* behave similarly: The proportion of responses favoring *het* is in both cases larger than for the nouns only listed with common gender. For this reason, established nouns that receive both common and neuter gender in Van Dale (such as *aquaplaning*) are coded as *established het*. Note that only four anglicisms in the dataset are listed with both common and neuter gender in the dictionary (see appendix 1).

to the borrowed noun. For example, *prestatie* is the translational equivalent of English *achievement*, whereas *gastenverblijf* is the translational equivalent of English *guesthouse* (which also contains a Dutch cognate, *huis*). To determine the most suitable translation, we use a translation dictionary (Van Dale Lexicografie 2006).

Rather than using the presence or absence of a cognate or a translation as a predictor, we instead rely on the gender listed for that cognate or for the translation of an anglicism. The gender of the cognate or translation is based on the Van Dale dictionary (Den Boon & Geeraerts 2008) and the online edition of the *Woordenlijst Nederlandse Taal* (Instituut voor Nederlandse Lexicologie & Nederlandse Taalunie 2005), which lists the official spelling and gender of a large number of Dutch words.

We expected that anglicisms would frequently adopt the gender of their Dutch lexical-semantic equivalent. Our initial analysis confirmed that for the purposes of gender assignment, an analogy is more likely to be drawn between the anglicism and its cognate than between the anglicism and its translation equivalent (see also Onysko et al. 2013). Therefore, we used a coding procedure for the lexical-semantic equivalence predictor that consists of two parts. As the first step, we checked whether an anglicism had a clear cognate. If so, we coded the gender of that cognate. If the anglicism did not have a clear cognate, we coded the gender of the translational equivalent. [12] For example,

---

[12] Initially, anglicisms without a clear cognate and without a translational equivalent (such as *aquaplaning*), on the one hand, and nouns with cognates or translational equivalents that can be either common or neuter (such as *busticket*: both *de ticket* and *het ticket* are possible), on the other hand, formed a separate category in the analysis. However, our preliminary analyses indicated that the first group of nouns (with no Dutch cognate or translation) behave similarly to the nouns with a common gender equivalent. Therefore, we collapsed these groups of nouns into one category, *de*. Since in this case there may be no conflict between the gender of the cognate/translation equivalent and the default gender (common gender, according to previous studies), we did not expect the gender of the Dutch equivalent to have a significant effect on the choice of the article. For two anglicisms in the dataset whose equivalents could be either common or neuter, our preliminary analyses indicated that they show the same behavior as the anglicisms with neuter gender equivalents. This is expected, because for these nouns, a conflict between the default gender and

skateboard is translated as a common noun *rol(schaats)plank* in the dictionary. However, it receives neuter gender as far as the predictor article of lexical-semantic equivalent is concerned: It occurs with the article *het* by analogy with *bord*, its neuter cognate. The anglicism *beach* does not have a clear cognate, so the neuter gender of its translation ('strand') is used for the purposes of article of lexical-semantic equivalent. Table 4 provides an overview of the distribution of anglicisms in the questionnaire with respect to the predictor article of lexical-semantic equivalent. In general, anglicisms with a common gender equivalent (article *de*) are much more frequent than anglicisms with a neuter equivalent (article *het*).

article of lexical-semantic equivalent	frequency in questionnaire	examples
de	133	*dancefloor* (Dutch cognate: *de dansvloer*) *engine* (Dutch translation: *de motor*)
het	42	*character* (Dutch cognate: *het karakter*), *sound* (Dutch translation: *het geluid*)

Table 4. Absolute frequencies of anglicisms
for the predictor article of lexical-semantic equivalent.

Our next predictor is *animacy of the referent of the anglicism*. A small group of native Dutch nouns receive a certain gender because of the semantic field to which they belong (see section 3.1; Haeseryn et al. 1997). Nouns with animate referents, including names for occupations and animals, are generally common (*de onderwijzer* 'the teacher', *de leeuw* 'the lion'), while names of sports, games, and metals are usually neuter (*het voetbal* 'the soccer', *het goud* 'the gold'). However, with the exception of nouns referring to sports and games, and nouns referring to

the equivalent's gender is possible. For this reason, we added these nouns to the group of anglicisms with a neuter equivalent.

persons, most loanwords do not belong to these semantic categories.[13] Therefore, we focus on the animacy of the anglicisms in the dataset to determine whether animate loanwords show a preference for the common gender, as is the case for the native Dutch nouns.[14]

An anglicism is coded as "animate" if the referent of the noun is a human or an animal (*believer*, *bulldog*). All other anglicisms are labeled "inanimate" (*busticket*, *container ship*). Table 5 provides an overview of the absolute frequencies of the anglicisms in the questionnaire with respect to the animacy predictor. Since native Dutch nouns with animate referents are usually common, we expected anglicisms with animate referents to occur frequently with the common gender article *de* as well.

animacy	frequency in questionnaire	examples
inanimate	148	*facelift*, *reality*
animate	27	*believer*, *nanny*

Table 5. Absolute frequencies of anglicisms for the predictor animacy.

The next set of predictors in our analysis concerns lectal features. We used personal information collected in the second part of the questionnaire to establish whether participants' sociolinguistic background plays a role in what gender they assign to anglicisms. More specifically, we wanted to find out whether the participants' gender, place of residence, parental educational level (as a proxy for the socio-

---

[13] Two anglicisms belonging to the semantic classes that play a role in the native gender system occur in Dutch: *indian summer* (name of a season) and *slang* (name of a language).

[14] Alternative coding procedures that can reveal whether semantic features have an effect on the gender of anglicisms were used as well. More specifically, the degree of individuation of the anglicism (animate, count/mass noun, concrete/abstract noun) was taken into account. Furthermore, we also coded each anglicism for the applicable semantic fields, according to the Historical Thesaurus of the online OED (HTOED; Oxford University Press 2013). However, these operationalizations do not reach significance in a multifactorial environment.

economic status of the participants), level of English proficiency, and knowledge of a local dialect correlate with their responses (see table 6).[15]

predictor	levels	# of respondents
gender of respondent	female	35
	male	10
region	central region of Flanders	24
	eastern region of Flanders	13
	western region of Flanders	8
parents' education	high	27
	low	18
English proficiency	good	25
	bad	20
knowledge of local dialect	no	26
	yes	19

Table 6. Overview of lectal variables.

Overall, the number of participants in the study is relatively low (N=45). This may explain why, as the results show, none of the lectal features reach significance in a multifactorial environment.

4.3. Methodology.
Previous research on gender of anglicisms in Dutch relies on raw frequencies or proportions to determine the relative importance of the

---

[15] More specifically, we measured the educational background of the parents by calculating the maximum of the scores of required educational background for the professions of both parents in the Standard Classification of Professions (*Standaard Beroepenclassificatie*) of the Dutch Central Bureau for Statistics (*Centraal Bureau voor de Statistiek* 2010). This classification ranges from 0 (jobs requiring the lowest level of education) to 8 (jobs requiring the highest level of education). If the maximum required educational level was higher than or equal to 6, the participant's parents were coded as having an "extensive" educational background; if the score was lower than 6, their educational background was coded as "not extensive".

factors that influence gender assignment. Moreover, to identify gender assignment rules, these studies generally use small corpora or lists of anglicisms collected from dictionaries (see, for instance, Geerts 1996). We aim to complement these studies by using inferential statistics to determine which factors have a significant influence on the choice of gender. More specifically, we use multiple mixed-effects logistic regression to model the effect of lectal and language-internal features on the binary response variable *article* (common *de* or neuter *het*).

A major advantage of a multifactorial regression model is that it can assess the impact and the significance of each of the predictors while taking into account the combined influence of all the variables in the model. Furthermore, this model allows for the inclusion of random effects. Using random effects is appropriate when a factor cannot be replicated, in the sense that the levels of the factors are not fixed and would differ if the experiment is repeated (Baayen 2008). For instance, the levels of the variable *gender* remain the same (*male*, *female*) every time the experiment is repeated. By contrast, the anglicisms and respondents in our questionnaire are sampled from a large pool of possible anglicisms and respondents, and are therefore highly unlikely to reoccur from experiment to experiment. Moreover, each respondent in the dataset chose an article for 175 anglicisms. Accordingly, the dataset contains some respondent-specific regularities: The responses of each respondent are probably correlated. The same holds for each of the anglicisms: Forty-five respondents chose an article for each of the English nouns, so the responses per anglicism are probably correlated as well. To cope with this type of respondent-specific and anglicism-specific variation, random factors can be included in a logistic regression model.

5. Determinants of Gender Assignment to Anglicisms in Dutch.
5.1. Predicting Anglicism Gender.
This section inquires into the factors that influence the gender chosen for anglicisms in Dutch. Overall, the participants in the study showed a preference for the common gender article *de* (see table 7).

article	absolute frequency	relative frequency
de	6,629	84.98%
het	1,172	15.02%
N/A	74	0%

Table 7. Distribution of response variable *article*.

Using a forward stepwise selection procedure, we built a mixed-effects logistic regression model. We examined the influence of all the predictors discussed above, namely, establishment, the article of the lexical-semantic equivalent, animacy of the referent, and the lectal features (the participants' gender, home region, parental educational level, level of English proficiency, and knowledge of a local dialect). Three predictors reach significance at the 0.05 level: the degree of establishment of the anglicism, the animacy of its referent, and the article of its lexical-semantic equivalent.[16] Interaction effects were taken into consideration, but they did not contribute enough to the explanatory power of the model to be included in the final model. Our model also contains random intercepts for two factors: *anglicism* and *respondent*. We checked whether by-subject random slopes could be added to the model, but the data do not support a model that is more complex than the model with two random intercepts. Diagnostics reveal a good fit of our model to the data.[17] The model performs well: It predicts 91.33% of

---

[16] First, a model containing only fixed effects was built to assess the power and goodness of fit of the model. Five predictors reached significance in this model, but two predictors lost significance when the random factors were added to the model.

[17] Initially, we confirmed that there were no harmful associations between all the predictors considered for the analysis. We assessed the goodness of fit of the fixed effects-only model in three ways: with an outlier analysis (no outliers were found), with a test for multicollinearity (all VIF-scores were below 2, with VIF-scores higher than 4 pointing to harmful associations between predictors), and by calculating the overdispersion parameter (Faraway 2006). This parameter is close to 1, which indicates that the residual deviance is approximately χ^2 distributed with the appropriate degrees of freedom. The diagnostics of the fixed effects-only model indicate that the power of the model is good. Pseudo R^2, a value ranging between 0 and 1, is 0.286, which indicates

the variants in our dataset correctly (compared to a baseline of 85% for a model that always chooses the most frequent variant *de*). The model's performance is also confirmed by the high C-value of 0.93 (a C-value of 0.8 or higher indicates that the model has predictive power).

Table 8 presents the output for the fixed effects in the mixed model. The predictors are presented in their relative order of importance: The article of the lexical-semantic equivalent and the degree of establishment of the anglicism have a significant influence on the choice between *de* and *het*. Animacy does not influence the alternation as much as the other two variables.

One reference level is chosen for each of the variables in the model. This level is included in the intercept, and so it does not receive separate values in the output. For instance, the reference level for article of lexical-semantic equivalent is *de*. The estimates, which are shown in the second column of the table, convey the direction and effect size of the impact of each predictor. These estimates for the levels of each predictor should be compared to the estimate for the intercept. The final column reports the p-value for each of the estimates (alpha level=0.05). The model predicts *het*, which means that positive estimates with a p-value smaller than 0.05 indicate that the odds of respondents choosing *het* are higher in comparison to the reference level, while significant negative estimates indicate that the probability of respondents choosing *de* is higher in comparison to the reference level. In other words, positive estimates indicate a higher probability of *het*, whereas negative estimates indicate a higher probability of *de*.

---

a decent fit for a logistic model. The C-value, which also ranges between 0 and 1, is 0.804, indicating predictability. With regard to the random effects structure of our model, we checked whether the Best Linear Unbiased Predictors (BLUPs) are normally distributed. Although a Shapiro-Wilk test reaches significance (p<0.01) for both of our random effects, which indicates that the BLUPs are not normally distributed, we also inspected the distribution of the BLUPs visually; we established that the distribution was good enough to permit drawing conclusions on the basis of our mixed model. We also checked whether including the fixed effects in our model would ensure a decrease in standard deviation around the BLUPs. Standard deviation decreases with 28.58% for the two random effects combined, which indicates that the model with the random factors is a better fit than a model that only includes fixed effects.

predictor	estimate	Std. Error	z-value	p-value
(Intercept)	-6.5453	0.5620	-11.646	0
article of lexical-semantic equivalent				
reference level: de				
het	1.9898	0.3436	5.791	0
establishment				
reference level: established *de*				
established *het*	3.2649	0.5933	5.503	0
non-established frequent	0.6445	0.3776	1.707	0.0879
non-established infrequent	1.2088	0.3725	3.245	0.0012
animacy				
reference level: animate				
inanimate	2.5290	0.5259	4.809	0

Table 8. Output for the fixed effects
in the mixed effects logistic regression model.

Table 8 shows that, not surprisingly, the gender of the Dutch cognate or translation equivalent often correlates with the gender assigned to an anglicism. More specifically, the chance of respondents using *het* with an anglicism is significantly higher when its Dutch lexical-semantic equivalent occurs with *het* than when it occurs with *de*. For example, only 1 out of 45 respondents used *de* for *skateboard*, which has a clear neuter cognate *het bord*, whereas for *input*, which has a common gender translation, *de invoer*, 44 out of 45 respondents selected *de*.

Moreover, table 8 confirms that the gender assigned to established anglicisms in the dictionary is a significant predictor of gender assignment. The probability of respondents using *het* is much higher if the anglicism is listed with *het* in the dictionary (for example, 31 out of 45 respondents use *het* for *entertainment*). In contrast, the probability of *het* is lower if the anglicism is listed as a common noun with *de* (for example, only 1 out of 45 respondents uses *het* for *knowhow*). Furthermore, table 8 also reveals significant differences between established loanwords listed with *de* and nonestablished anglicisms: The probability of *de* is lower for nonestablished anglicisms, especially if they are infrequent. This may be due to a strong correlation between established nouns that are listed with *de* and usage of the common

gender determiner. For nonestablished anglicisms, such a correlation does not exist.

The estimates for the third predictor in the model, that is, animacy, indicate that respondents are more inclined to use the neuter article *het* for inanimate nouns. This is in accordance with the general agreement system of Dutch, as described in Haeseryn et al. 1997: Nouns referring to humans and animals are generally common.

In sum, the analysis indicates that anglicisms generally occur with the common gender article *de* in Dutch, but the likelihood of *het* increases in certain cases. More specifically, when the anglicism has a neuter gender cognate or translation equivalent, when the anglicism is established and listed as neuter in the dictionary or when it is nonestablished and infrequent, and when the referent of the anglicism is inanimate, the likelihood of *het* is significantly higher.

The output for the random effects included in the model, namely, anglicism and respondent, offers some further insight into the structure of the variation found in the data set. Table 9 provides an overview of the variance and standard deviation associated with the random factors (both adjustments to the intercept) in the model. The reported variances reflect the importance of the random effects.

Examining the random intercepts reveals that the highest positive adjustments are made for the anglicisms *character* (4.55) and *shoppingcenter* (3.97). This means that, all other things being equal, the neuter gender article *het* is used for these nouns exceptionally frequently. The highest negative adjustments are associated with *baseballbat* (-3.32) and *countdown* (-3.20). For these nouns, the common gender article *de* is favored. Some by-subject variation occurs in the dataset as well. The intercept adjustments indicate that two respondents, with ID's 13 (2.22) and 36 (2.10), select the neuter gender article *het* much more frequently than the other respondents, all other things held constant. Both of these respondents are 21-year-old females, but they do not have any other socio-economic features in common. The highest negative intercept adjustments are made for respondents 3 (-1.15) and 17 (-0.75). However, the difference with the other subjects is not very large.

random effect (intercept adjustments)	variance	standard deviation
anglicism	2.8546	1.6895
respondent	0.5385	0.7339

Table 9. Variance and standard deviation
of the random factors in the regression model.

Overall, Table 9 reveals that most of the variation is accounted for by intercept adjustments per anglicism, although differences between respondents are important as well.

5.2. (Dis)agreement Among Respondents: Analysis and Results.

In this section, we examine the extent to which the respondents agree or disagree regarding the article they assign to an anglicism. As the discussion of the random factors in the regression model indicated, by-item intercept adjustments explain most of the variation. By-subject intercept adjustments contribute significantly to the model as well, which means that the respondents do not always agree about the gender they assign to a particular anglicism.
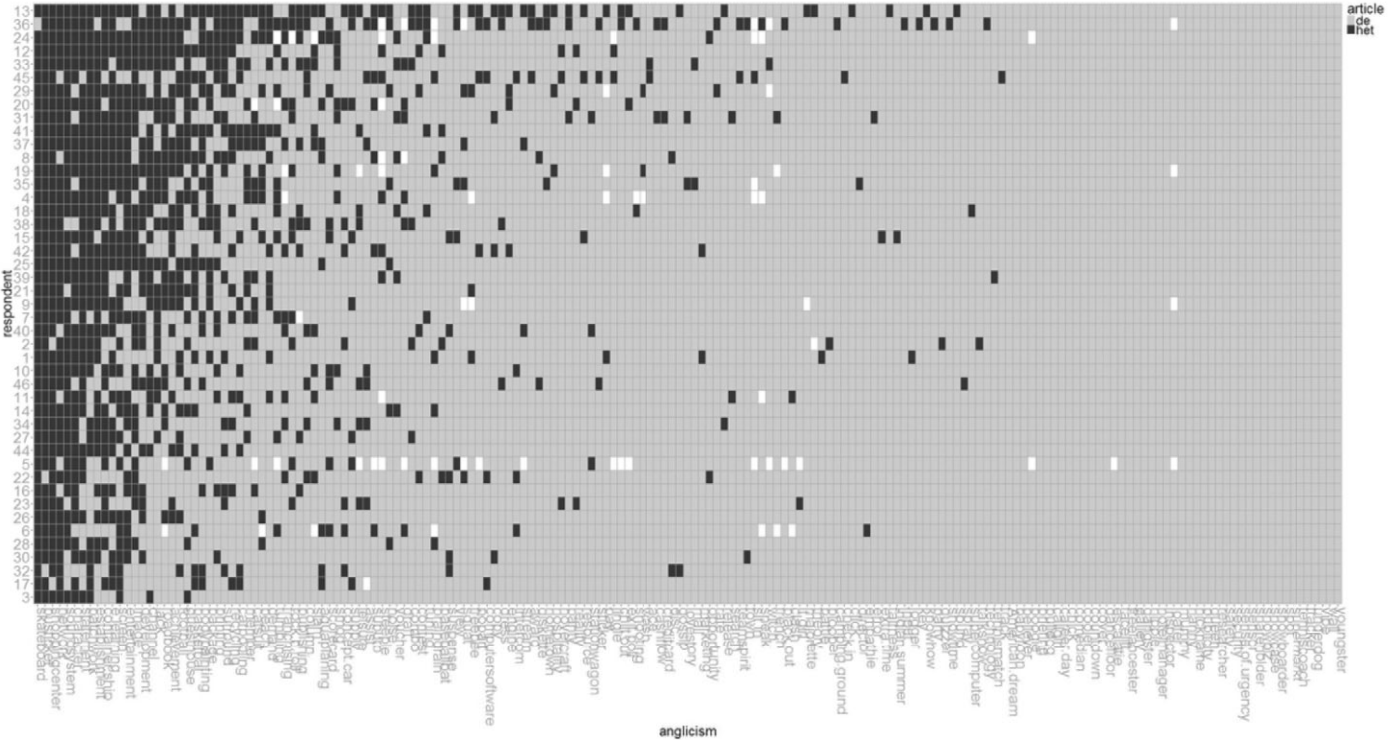
Figure 2. Plot of article chosen per anglicism and per respondent.

Figure 2 provides a visualization of the variation in our dataset per anglicism and per respondent. In the plot, a grey box indicates that the respondent selected *de*; a black box indicates that the respondent chose *het*, while a white box represents a missing value. For example, respondent 12 chose *het* for the anglicism *jingle*, while respondent 22 selected *de* for this noun. Respondents 5 and 24 did not make any selection. For *computersoftware*, respondent 17 selected *het*, while respondent 27 chose *de*. The anglicisms are ordered (on the x-axis) by decreasing proportion of *het* responses from left to right. The respondents are ordered (on the y-axis) by decreasing proportion of *het* responses from top to bottom.

The plot reveals that the proportion of neuter (or common) gender responses differs greatly per respondent and per anglicism. Respondent 3 chooses *het* the least (for 11 out of 145 anglicisms), while respondent 13 chooses *het* most frequently (for 66 out of 145 anglicisms). All respondents but one select *het* for *skateboard*, while for a relatively high number of anglicisms, on the right-hand side of the plot (that is, *youngster*, *wall*, *vibe*, etc.), all respondents choose the same article, *de*. Interestingly, there is not one anglicism in the dataset for which all respondents select *het*.

Furthermore, the plot corroborates the finding that within-respondent variation occurs in the dataset as well: Our respondents did not systematically opt for the same article for two anglicisms with similar properties. For instance, respondent 36 assigns neuter gender to *teamspirit*, an inanimate, nonestablished frequent anglicism, with a common gender Dutch cognate (*spirit*); however, the same respondent assigns common gender to a similar anglicism, (*American*) *dream*—also inanimate, nonestablished, frequent, and with a common gender Dutch cognate (*droom*).[18]

Crucially, figure 2 also reveals variation in the amount of disagreement among the respondents in the dataset per anglicism. For some anglicisms, such as *skateboard*, *youngster*, *wall*, and *vibe*, all or

---

[18] Some of the anglicisms appeared with an English modifier in the questionnaire because the nouns only appeared in certain collocations in the sources that were consulted (see section 4): *American **dream***, *camping **ground***, *career **day***, *concept **car***, *container **ship***, *indian **summer***, ***sense** of urgency*. The article is assigned to the head nouns of each of these multi-word expressions.

most of the respondents agree about the appropriate gender, but for other anglicisms, there is much more variation. For the fitness term, *squat*, for instance, about 85% of the respondents (38 out of 45) choose the same article, *de*; for *franchising*, about 71% of the respondents (32 out of 45) agreed on *de*; for *achievement*, there is very little agreement: 22 respondents selected *de*, and 23 respondents selected *het*.

By relying on the predicted probabilities of the regression model, we can determine whether the amount of variation with respect to the anglicisms is influenced by the predictors discussed in section 4.2. More specifically, for each combination of the variables included in the questionnaire, we predicted the probability of the respondents selecting *het*. Disagreement is highest if this probability is close to 50%; if it is close to 0% or 100%, respondents show more agreement (they either all choose *de*, or all choose *het*).

Table 10 presents an overview of all the combinations of the levels of the predictors in the dataset. Columns 2, 3, and 4 show the level used per predictor. Columns 5 and 6 show the mean predicted probability of *het* and the standard deviation for that combination of levels. Columns 7 and 8 indicate how many records are available for the combination of predictors.

	article of equivalent	establishment	animacy	mean predicted probability	standard deviation predicted probability	total # of records	# of anglicisms in questionnaire
1	*de*	established *de*	animate	0.0024	0.0044	495	11
2	*het*	established *de*	animate	0.011	0.0151	135	3
3	*de*	nonestablished frequent	animate	0.0062	0.0097	270	6
4	*de*	nonestablished infrequent	animate	0.0087	0.0121	225	5
5	*het*	nonestablished infrequent	animate	0.5288	0.2659	90	2
6	*de*	established *de*	inanimate	0.0639	0.0932	1980	44
7	*het*	established *de*	inanimate	0.0736	0.1082	360	8
8	*de*	established *het*	inanimate	0.4687	0.3142	180	4
9	*het*	established *het*	inanimate	0.6475	0.2752	360	8
10	*de*	nonestablished frequent	inanimate	0.0783	0.1240	1440	32
11	*het*	nonestablished frequent	inanimate	0.3862	0.3263	495	11
12	*de*	nonestablished infrequent	inanimate	0.1146	0.1691	1395	31
13	*het*	nonestablished infrequent	inanimate	0.4111	0.2724	450	10

Table 10. Predicted probabilities (for *het*) of each combination of predictors present in the dataset.

The table shows that the probability of using *het* is never higher than 65%. Furthermore, the greatest variability in the participants'

responses is observed when the mean predicted probability of *het* is close to 50%. The table indicates that this happens in case of a conflict between different factors that play a role in gender assignment. This is apparent in three sets of cases. First, the mean predicted probability of *het* is close to 50% in the case of 12 inanimate, established anglicisms listed with *het* (rows 8 and 9 in table 10). In this case, the article of their Dutch equivalents has only a minor influence. Row 9 shows the predicted probability of using *het* for established anglicisms for which there is no conflict between the neuter listed gender and the neuter gender of a Dutch equivalent. It is noteworthy that even if an established anglicism is listed as neuter and has a neuter Dutch equivalent, the probability of using *het* is still relatively low (64.75%). This is especially evident if one compares the predicted values for such neuter anglicisms to the predicted values for the inanimate established anglicisms in row 6: These anglicisms are listed with *de* and have a common gender Dutch equivalent. Most of the time, the respondents agree about their gender, as the probability of *het* is only 6.39%. This difference can be explained as follows: When properties that favor the neuter article are available, a conflict arises between these properties and the default article for anglicisms, *de*. Previous studies have argued that anglicisms receive the common definite article *de*, unless there is a reason for using neuter gender *het* (see section 2.2). However, our data indicate that respondents do not categorically select neuter gender in this case; instead, independent reasons for using neuter gender *het* causes greater disagreement among speakers.

Second, disagreement among respondents is relatively high in the case of nonestablished anglicisms that have a neuter gender Dutch equivalent. If these nouns have an animate referent, the disagreement among respondents is the greatest (52.88% probability of choosing *het*). However, note that only two nouns in the questionnaire belong to this group. These findings can be explained by the general semantic tendency in Dutch: Nouns that refer to people and animals are usually assigned common gender. As a result, there is a conflict between the common gender often used for nouns with an animate referent and the neuter gender of the Dutch equivalent.

Third, it is unexpected that the proportion of *het* is relatively low with the inanimate nonestablished nouns (38.62% and 41.11%). The 21 nouns that belong to this group do not display any features that favor

common gender. At the same time, one salient feature that favors neuter gender—that is, inanimacy—is readily available. Nevertheless, many respondents prefer the default common gender article for this group of nouns.

In sum, these results seem to contradict the claim often made in the literature that gender variation (that is, disagreement in the speech community with respect to gender assignment) only concerns nonestablished nouns. In our data, this variation concerns established nouns as well, if they do not have the default gender listed in the dictionary. Consequently, establishment in the speech community does not serve as the sole explanation for an increase in homogeneity in our dataset. Furthermore, the analysis indicates that in the case of conflict between multiple factors that influence the gender of an anglicism— such as a neuter lexical-semantic equivalent versus the default common gender—homogeneity decreases. In other words, conflicting factors seem to lead to more heterogeneity in the speech community.

6. Discussion.
6.1. Previous Research and Current Findings.
In this paper, we focused on the gender of anglicisms in Dutch. We have reviewed previous studies noting their treatment of two issues: how gender is assigned to loanwords (section 2.1) and variability in gender assignment to loanwords (section 2.2). With respect to the first issue, it has been generally assumed that loanwords are subject to the gender-assignment rules of the host language. Furthermore, it has been proposed that loanwords can be assigned gender by semantic analogy. Finally, it has been proposed that loanwords can also be assigned the so-called default gender of the host language. Previous research suggests that in Dutch, anglicisms receive common gender by default (*de*), unless there is a reason for them to receive neuter gender (*het*). Such reasons can be semantic, morphological, or analogy-based. However, with the exception of Geerts (1996), who discusses a hierarchy of gender assigning rules, scholars do not elaborate on the interaction between these determinants of anglicism gender.

With respect to the second issue—that is, variability in gender assignment—scholars have argued that variability or the lack thereof is related to (non-)establishment (but see Rothe 2014): Only those loanwords show variable gender that have not been established in the

host language; once a loanword is established, the speakers generally agree about its gender.

This paper contributes to the debate through introducing novel research tools. We utilized inferential statistics to investigate gender assignment to anglicisms in Dutch. First, we analyzed the factors relevant for gender assignment to English loanwords in Dutch. This analysis yielded a model whose predictions were then used to assess the amount of (dis)agreement among the respondents when it comes to the gender of individual anglicisms. Furthermore, since the data included both established and nonestablished borrowings, we were able to verify whether the degree of establishment of a particular anglicism affects the amount of (dis)agreement among the speakers concerning its gender.

With respect to the factors underlying gender assignment to anglicisms, our findings are as follows. Our analysis confirms the prevalence of the default common gender article *de*: Only 15.02% of our data contain the neuter gender article *het*. Furthermore, the regression model provides evidence for the importance of the Dutch gender system: Anglicisms with an animate referent generally appear with the common gender article *de*, just like native Dutch nouns. In addition, the model suggests that anglicisms frequently receive the article of their Dutch cognate or translation equivalent. Finally, the degree of establishment of an anglicism also plays an important role in the respondents' choice of an article.

Let us now turn to the issue of variability in gender assignment. The random factors in our regression model reveal inter-respondent variation in gender assignment to nominal anglicisms. A visualization of the results shows that the degree of variation varies from anglicism to anglicism. In addition, intra-respondent variation occurs as well: Language users do not systematically choose a particular gender for anglicisms with similar properties. The regression model also reveals that disagreement among the speakers is particularly high in case of (i) established anglicisms with an inanimate referent listed with *het*, and (ii) nonestablished anglicisms with a neuter gender Dutch equivalent. In other words, the model shows less homogeneity in the speech community if there is a conflict between the default article *de* and one or more factors that favor neuter gender.

It should be noted that previous studies consider the default article a last resort (for example, Geerts 1996, Hock & Joseph 1996:266–269):

An anglicism would be assigned the default gender only if all other gender-assignment rules fail. However, our analysis indicates that this is not the case: Speakers are more likely to disagree about the gender of an anglicism that qualifies for the nondefault neuter gender; there is generally more consensus about the gender of anglicisms with a common gender Dutch equivalent. This is because language users do not automatically opt for neuter gender in the presence of a neuter gender favoring factor. Even when an anglicism has a very clear neuter gender cognate (see the *skateboard* example above), not all the respondents automatically assign neuter gender to the anglicism. This means that the last resort approach cannot account for the gender assignment pattern in Dutch.

Our results also contrast with Geerts (1996). He assumes that in the case of a conflict between the default article *de* and the presence of a factor that favors the neuter article *het*, respondents unanimously choose the neuter article. Interestingly, our results indicate that in the case of such a conflict, there is more variability: Respondents find it harder rather than easier to choose an article for the anglicism. Furthermore, in our dataset, respondents never categorically choose *het* for any anglicism.

Moreover, Geerts (1996) argues that the rules applicable to anglicisms can be arranged along a clear hierarchy, as shown in 1.[19] When multiple rules apply, some rules take precedence over others.

(1)   cognate   >   prototypical   >   suffix   >   gender of
      gender   gender of   gender   translational
                     semantic field   equivalent

However, our dataset contains several anglicisms whose gender cannot be explained by this hierarchy. The anglicism *member*, for instance, receives neuter gender by virtue of having a neuter lexical-semantic equivalent (its neuter translation *lid*); at the same time, as a noun with an animate referent, it should acquire common gender. Furthermore, *member* is nonestablished and infrequent. It seems that, due to the conflict between the multiple factors that apply to the noun, and due to

---

[19] According to Geerts 1996:145, the gender of the translational equivalent is less crucial than the three other factors.

the availability of a default gender for anglicisms, our respondents disagree considerably on the gender of *member*: 31 respondents use *de*, while 14 respondents select *het*.

Finally, by explicitly including both established and nonestablished anglicisms in our questionnaire, we were able to examine whether establishment of an anglicism in the speech community reduces disagreement on its gender. In contrast with what has been argued in most other studies on gender assignment to anglicisms, our results indicate that the amount of variation does not depend solely on the degree of establishment of anglicisms. Our analysis shows that variation also depends on the interplay of other factors that influence gender assignment: Established anglicisms listed in the dictionary as neuter, for instance, show a high amount of variation as well, in comparison to established anglicisms listed with common gender.

6.2. *Issues for Further Research.*

While we were able to attenuate some hypotheses formulated in previous research by showing that language users do not necessarily agree about the gender of an anglicism, our study has some shortcomings that should be addressed in further research. First, the dataset only contains responses of 45 participants. Moreover, as all respondents were university students, follow-up research should include a higher number of speakers from a more stratified sample of the population. Even though we used background information on each respondent to operationalize some lectal features as independent variables, these features did not reach significance in a multivariate environment. Using a larger, more stratified sample might uncover respondent-specific features that impact gender assignment to loanwords as well.

Second, previous studies indicate that the gender associated with the suffix of an anglicism plays a role in gender assignment as well. However, we were not able to include this predictor in our regression model because of the low number of recently borrowed anglicisms with a suffix generally associated with neuter gender in Dutch. Exploratory analyses confirm that this predictor influences the gender of anglicisms: The proportion of neuter gender articles is significantly higher in anglicisms with suffix *-ment*, which is associated with neuter gender,

than in anglicisms with a suffix associated with common gender. However, follow-up studies are needed to corroborate this result.

Third, previous studies have argued that anglicisms belonging to particular semantic fields, such as breeds of dogs, drinks, collectives, sports, and substances, often receive a particular gender (see section 2.2). This claim is not entirely supported by our study. The only evidence we obtained for the relevance of semantic features is that anglicisms with an animate referent appear significantly more often with the common gender article *de* than nouns with an inanimate referent. Note, however, that most of the semantic fields discussed in previous studies could not be included in our analysis because of the low number of anglicisms belonging to these fields. Therefore, an interesting extension to this study might focus on a different host language with a semantic gender system.

Fourth, the Van Dale dictionary, an implicitly normative dictionary of Dutch, was used in this study to determine whether an anglicism is established or not. Words are included in this dictionary if they are found regularly and for an extensive period of time in the readily available language of the speech-making community (that is, writers, teachers, scientists, journalists, etc.; Den Boon & Geeraerts 2008). The more recent editions of the dictionary rely more heavily on digital texts to determine which lexical items should be included. For this reason, whether or not an anglicism is listed in the dictionary is often determined by its frequency. However, in some cases, frequency may not be on a par with the acceptedness or normative status of the word.

To remedy this shortcoming, follow-up studies may want to adopt an alternative, usage-based approach for determining whether a noun is established in the speech community. Factors that influence the success of anglicisms in Dutch have been examined using the entire onomasiological profile of the concept in large corpora—that is, the anglicism and all its (native) counterparts that express the same concept (Zenner et al. 2012, Zenner et al. 2014). The degree of establishment of an anglicism can then be quantified by taking into account its relative weight vis-à-vis the frequency of its (native) synonymous expressions. A second method that can be envisaged to determine the establishment of an anglicism relies on the degree to which a noun is adapted to the host language. Orthographic adaptation, for instance, can serve as an indication of a relatively high degree of establishment in the host

language. Finally, attitudinal data may also provide information about the acceptedness of non-native material in the host language.

Further extensions of this study are also possible. The quantitative methodology applied in this paper offers a new way of investigating the adaptation of loanwords from gendered source languages. By taking into account both the native gender system and the homogeneity in the responses one may discover further criteria that do not apply to borrowings from a nongendered language, such as English. Alternatively, one could concentrate on host languages with more rule-based gender systems (as the gender system of Dutch is, to a large extent, arbitrary). This approach could reveal whether variability in the amount of (dis)agreement in the speech community is also related to the arbitrariness of the Dutch gender system.

Moreover, the preponderance of the common gender article *de* is expected on the basis of previous studies on Dutch. However, few scholars have asked why neuter gender was conventionalized for anglicisms in the first place; why not rely on the default common gender alone for all anglicisms? One exception is Geerts (1970), who argues that Dutch undergoes a general process of deneutralization (*ont-het-ting* 'de-*het*-ing'). This explanation was first offered by Van Haeringen (1951) for concrete mass nouns. Geerts proposes that the neuter gender for nouns in Dutch is a relatively closed category: Nouns only receive this gender when very compelling reasons apply. For anglicisms, such compelling reasons are usually not available: They often refer to novel objects or concepts, and they frequently lack the formal properties associated with neuter gender. However, a systematic diachronic quantitative study is necessary to corroborate this explanation.

7. Conclusion.

Our paper offers valuable new insights and helps refine some of the hypotheses that have been put forward before. Methodologically, our multivariate approach enables us to demonstrate which factors influence the assignment of gender to anglicisms in Dutch. Furthermore, this methodology allows us to show that the degree of homogeneity in the speech community does not solely depend on the degree of establishment of an anglicism. Gender assignment to anglicisms in Dutch is not a categorical process, in contrast with what has been assumed in previous

studies. On the theoretical level, we find clear evidence for the status of the common gender article *de* as the default article for anglicisms in Dutch. Regarding variation in the amount of agreement about the gender of an anglicism, the analysis indicates that homogeneity decreases when multiple rules apply. In sum, our analysis shows how the interplay between several factors influences both gender assignment to anglicisms in Dutch and disagreement within the speech community.

APPENDIX
List of Anglicisms

1. Established nouns

anglicism	gender in Van Dale	gender in Van Dale (binary)	establishment
ace	both	*het*	established *de*
aquaplaning	none	*de*	established *het*
assist	*de*	*de*	established *de*
babyfoon	*de*	*de*	established *de*
baseballbat	*het*	*het*	established *het*
basketbal	*het*	*het*	established *het*
black-out	*de*	*de*	established *de*
blooper	*de*	*de*	established *de*
braindrain	*none*	*de*	established *de*
briefing	*de*	*de*	established *de*
bulldog	*de*	*de*	established *de*
busticket	both	*het*	established *het*
callgirl	*de*	*de*	established *de*
carpooling	*het*	*het*	established *de*
cashflow	*de*	*de*	established *de*
chick	*het*	*het*	established *de*
clash	*het*	*het*	established *de*
click	*de*	*de*	established *de*
column	*de*	*de*	established *de*

anglicism	gender in Van Dale	gender in Van Dale (binary)	establishment
combo	*de*	*de*	established *het*
computersoftware	*de*	*de*	established *de*
cover	none	*de*	established *de*
creditcard	*de*	*de*	established *de*
disc	*de*	*de*	established *de*
diskette	*de*	*de*	established *de*
empathie	*het*	*het*	established *de*
entertainment	*het*	*het*	established *het*
escalatie	none	*de*	established *de*
establishment	*de*	*de*	established *het*
extra-time	*de*	*de*	established *de*
facelift	none	*de*	established *de*
franchising	*de*	*de*	established *de*
freelancester	none	*de*	established *de*
gangster	*het*	*het*	established *de*
gossip	*de*	*de*	established *de*
hippie	none	*de*	established *de*
hotelmanager	*de*	*de*	established *de*
hovercraft	*het*	*het*	established *de*
image	*het*	*het*	established *het*
input	*de*	*de*	established *de*
intercom	none	*de*	established *de*
jack	*het*	*het*	established *het*
jingle	*de*	*de*	established *de*
jogger	*de*	*de*	established *de*
knowhow	*de*	*de*	established *de*
loafer	*de*	*de*	established *de*
lovestory	*de*	*de*	established *de*
majorette	*het*	*het*	established *de*
marketing	*de*	*de*	established *de*

anglicism	gender in Van Dale	gender in Van Dale (binary)	establishment
mummy	*de*	*de*	established *de*
panty	*de*	*de*	established *de*
party	*de*	*de*	established *de*
patchwork	*het*	*het*	established *het*
popart	none	*de*	established *de*
recycling	*de*	*de*	established *de*
release	*de*	*de*	established *de*
researcher	*de*	*de*	established *de*
rock'n'roll	*de*	*de*	established *de*
sensor	none	*de*	established *de*
sit-in	*de*	*de*	established *de*
skateboard	*het*	*het*	established *het*
smog	*de*	*de*	established *de*
snowboarder	none	*de*	established *de*
sound	*het*	*het*	established *de*
spot	*het*	*het*	established *de*
spray	*de*	*de*	established *de*
statement	*de*	*de*	established *het*
stationwagon	*de*	*de*	established *de*
sticker	none	*de*	established *de*
supercomputer	*de*	*de*	established *de*
supermarkt	*de*	*de*	established *de*
suspense	*de*	*de*	established *de*
talkshow	*de*	*de*	established *de*
tenniscoach	*de*	*de*	established *de*
tennismatch	*de*	*de*	established *de*
track	*het*	*het*	established *de*
underdog	*de*	*de*	established *de*
voucher	*de*	*de*	established *de*

2. Nonestablished nouns

anglicism	corpus frequency	frequency (categorical)
achievement	168	infrequent
American dream	1743	frequent
award	2515	frequent
battle	695	frequent
beach	1111	frequent
believer	184	infrequent
bodypainting	281	infrequent
boogie	566	frequent
camping ground	767	frequent
career day	4240	frequent
catsuit	100	infrequent
channel	134	infrequent
character	140	infrequent
check-in	811	frequent
chill-out	200	infrequent
comedian	1905	frequent
concept car	1656	frequent
container ship	151	infrequent
contest	326	frequent
cooler	243	infrequent
copy	158	infrequent
countdown	160	infrequent
dancefloor	358	frequent
demand	369	frequent
development	974	frequent
director	3704	frequent
draft	304	frequent
dress	366	frequent
e-learning	212	infrequent

anglicism	corpus frequency	frequency (categorical)
engine	212	infrequent
error	285	infrequent
exit-poll	211	infrequent
fairness	198	infrequent
football	1313	frequent
freebee	133	infrequent
gallery	142	infrequent
guesthouse	188	infrequent
hometown	117	infrequent
hospitality	111	infrequent
indian summer	1682	frequent
inspector	228	infrequent
it-girl	2222	frequent
key	889	frequent
leisure	237	infrequent
mall	313	frequent
member	150	infrequent
nanny	375	frequent
network	512	frequent
nickname	134	infrequent
noise	777	frequent
opportunity	191	infrequent
package	185	infrequent
pay	744	frequent
printing	145	infrequent
publicity	115	infrequent
publishing	145	infrequent
queen	1042	frequent
quizzer	107	infrequent
rate	627	frequent
reality	1529	frequent

anglicism	corpus frequency	frequency (categorical)
resource	252	infrequent
roadbook	384	frequent
safety	514	frequent
scorecard	169	infrequent
screen	1396	frequent
search	476	frequent
security	1722	frequent
sense of urgency	604	frequent
shareholder	148	infrequent
shooting	384	frequent
shoppingcenter	885	frequent
showbiz	110	infrequent
showcase	407	frequent
showtime	157	infrequent
slam	450	frequent
sneak	111	infrequent
soundsystem	209	infrequent
squat	100	infrequent
steeple	3892	frequent
stream	205	infrequent
suburb	143	infrequent
supply	239	infrequent
surround	267	infrequent
teamspirit	1240	frequent
technology	542	frequent
theory	290	infrequent
town	1171	frequent
tracker	340	frequent
trench	107	infrequent
tumbling	699	frequent
tuning	342	frequent

anglicism	corpus frequency	frequency (categorical)
vibe	293	infrequent
view	586	frequent
wall	510	frequent
wash	191	infrequent
witch	184	infrequent
youngster	944	frequent

REFERENCES

Audring, Jenny. 2009. *Reinventing pronoun gender*. Utrecht: LOT.

Baayen, Harald R. 2008. *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Berteloot, Armand, & Nicoline Van der Sijs. 2003. Dutch. *English in Europe*, ed. by Manfred Görlach, 37–56. Oxford, NY: Oxford University Press.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, & Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education.

Budzhak-Jones, Svitlana. 1997. Quantitative analysis of gender assignment in mono/bilingual discourse. *Journal of Quantitative Linguistics* 4. 67–91.

Callies, Marcus, Alexander Onysko, & Eva Ogiermann. 2012. Investigating gender variation of English loanwords in German. *The anglicization of European lexis*, ed. by Cristiano Furiassi, Virginia Pulcini, & Félix Rodríguez González, 65–89. Amsterdam: John Benjamins.

Chirsheva, Galina. 2009. Gender in Russian-English code-switching. *International Journal of Bilingualism* 13. 63–90.

Clyne, Michael. 1992. *Pluricentric languages: Differing norms in different nations*. Berlin: Mouton de Gruyter.

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2014. Gender typology. *The expression of gender*, ed. by Greville G. Corbett, 87–130. Berlin: Mouton de Gruyter.

Cruz Cabanillas, Isabel de la, Cristina Tejedor Martínez, Mercedes Díez Prados, & Esperanza Cerdá Redondo. 2007. English loanwords in Spanish computer language. *English for Specific Purposes* 26. 52–78.

Faraway, Julian J. 2006. *Extending the linear model with R. Generalized linear, mixed effects and nonparametric regression models*. Boca Raton: Chapman & Hall.

Geerts, Guido. 1970. De nominale klassifikatie van ontleningen. *De Nieuwe Taalgids* 63. 43–53.

Geerts, Guido. 1996. De genusbepalende eigenschappen van Engelse leenwoorden in het Nederlands. *Verslagen en Mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde* 106. 137–146.

Haeringen, Coenraad B. Van. 1951. Genusverandering bij stofnamen. *De Nieuwe Taalgids* 44. 7–14.

Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jaap De Rooij, & Maarten C. Van Den Toorn. 1997. *Algemene Nederlandse Spraakkunst*. Groningen: Martinus Nijhoff uitgevers.

Hamans, Camiel. 2009. Het geslacht van leenwoorden. *Die tand van die tyd. Opstelle opgedra aan Jac Conradie*, ed. by Willie Burger & Marné Pienaar, 24–25. Stellenbosch: Sun Media.

Haugen, Einar. 1969. *The Norwegian language in America: A study in bilingual behavior*. Bloomington, IN: Indiana University Press.

Hock, Hans H., & Brian D. Joseph. 1996. *Language history, language change, and language relationship: An introduction to historical and comparative linguistics*. Berlin: Mouton de Gruyter.

Koenen, Liesbeth, & Rik Smits. 1992. *Peptalk. De Engelse woordenschat van het Nederlands*. Amsterdam: Nijgh & Van Ditmar.

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

Muysken, Pieter. 2000. *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.

Onysko, Alexander. 2007. *Anglicisms in German: Borrowing, lexical productivity, and written codeswitching*. Berlin: de Gruyter.

Onysko, Alexander, Marcus Callies, & Eva Ogiermann. 2013. Gender variation of anglicisms in German: The influence of cognitive factors and regional varieties. *Poznan Studies in Contemporary Linguistics* 49. 103–136.

Poplack, Shana, Alicia Pousada, & David Sankoff. 1982. Competing influences on gender assignment: Variable process, stable outcome. *Lingua* 57. 1–28.

Posthumus, Jan. 1996. Het woordgeslacht van Engelse leenwoorden. *Onze Taal* 65. 279.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. New York, NY: Longman.

Rothe, Astrid. 2014. On the variation of gender in nominal language mixings. *New perspectives on lexical borrowing: Onomasiological, methodological and phraseological innovations*, ed. by Eline Zenner & Gitte Kristiansen, 191–223. Boston, MA: Mouton De Gruyter.

Schenck, Alfons. 1985. *Genusbepalende eigenschappen van Engelse leenwoorden in het Nederlands*. Leuven, Belgium: KU Leuven MA thesis.

Smead, Robert N. 2000. On the assignment of gender to Chicano Anglicisms: Processes and results. *The Bilingual Review* 23. 277–297.

Thornton, Anna M. 2009. Constraining gender assignment rules. *Language Sciences* 31. 14–32.

Verhoeven, Gerard, & Frank Jansen. 1996. Het woordgeslacht van Engelse leenwoorden. *Onze Taal* 65. 156–157.

Violin-Wigent, Anne. 2006. Gender assignment to nouns codeswitched into French: Observations and explanations. *International Journal of Bilingualism* 10. 253–276.

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

Weinreich, Uriel. 1968. *Languages in contact: Findings and problems*. The Hague: Mouton de Gruyter.

Zenner, Eline. 2013. *Cognitive contact linguistics. The macro, meso and micro influence of English on Dutch*. Leuven, Belgium: KU Leuven dissertation.

Zenner, Eline, Dirk Speelman, & Dirk Geeraerts. 2012. Cognitive sociolinguistics meets loanword research: Measuring variation in the success of anglicisms in Dutch. *Cognitive Linguistics* 23. 749–792.

Zenner, Eline, Dirk Speelman, & Dirk Geeraerts. 2014. Core vocabulary, borrowability and entrenchment: A usage-based onomasiological approach. *Diachronica* 31. 74–105.

Dictionaries and Corpora

Boon, Ton Den, & Dirk Geeraerts. 2008. *Van Dale Groot Woordenboek van de Nederlandse Taal*. Utrecht: Van Dale Lexicografie. 14th edition.

Centraal Bureau voor de Statistiek [Dutch Central Bureau for Statistics]. 2010. *Standaard Beroepenclassificatie* [Standard Classification of Professions]. Available at *https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/onderwijs%20en%20beroepen/beroepenclassificatie--isco-en-sbc--*, accessed on 2 April, 2015.

Het Groene Boekje. Woordenlijst Nederlandse Taal. 2005. Instituut voor Nederlandse Lexicologie & Nederlandse Taalunie. Utrecht: Van Dale. Available at http://www.woordenlijst.org, accessed on 2 April, 2015.

Oxford English Dictionary. 2013. Oxford: Oxford University Press. Available at http://www.oed.com/, accessed on 2 April, 2015.

R: A language and environment for statistical computing. 2013. R Development Core Team. Vienna, Austria. Available at https://www.R-project.org/, accessed on 20 October, 2016.

Van Dale Groot woordenboek Engels-Nederlands. 2006. Utrecht: Van Dale Lexicografie.

QLVL, KU Leuven
Department of Linguistics
Blijde-Inkomststraat 21
PO box 3308
3000 Leuven
Belgium
[karlien.franco@kuleuven.be]
[eline.zenner@kuleuven.be]
[dirk.speelman@kuleuven.be]