

Extravagant “fake” morphemes in Dutch

Morphological productivity, semantic profiles and categorical flexibility

Abstract

Dutch features several morphemes with “privative” semantics that occur as left-hand members in compounds (e.g. *imitatieleer* ‘imitation leather’, *kunstgras* ‘artificial grass’, *nepjuwelen* ‘fake jewels’). Some of these “fake” morphemes display great categorical flexibility and innovative adjectival uses. *Nep*, for instance, is synchronically attested as an inflected adjective (e.g. *neppe cupcake* ‘fake cupcake’). In this paper, we combine an extensive corpus study of eight Dutch “fake” morphemes with statistical methods in distributional semantics and collexeme analysis in order to compare their semantic and morphological properties and to find out which factors are the driving forces behind their exceptional “extravagant” morphological behaviour. Our analyses show that debonding and adjectival reanalysis are triggered by an interplay of two factors, i.e. type frequency and semantic coherence, which allow us to range the eight morphemes on a cline from more schematic to more substantive “fake” constructions.

Keywords: compounding; privative semantics; productivity; debonding; Semantic Vector Spaces; multiple distinctive collexeme analysis; Dutch

1. Introduction

In this study, we present a corpus-based statistical analysis of eight Dutch “fake” morphemes that display, to a more or lesser degree, “extravagant” morphological properties. We combine

an extensive corpus study with statistical methods in distributional semantics and collexeme analysis in order to compare the semantic and morphological properties of these morphemes and to find out which factors are the driving forces behind the exceptional morphological behaviour of some of these “fake” morphemes.

Let us first define what we understand by “fake” morphemes. We adopt a broad definition in order to include all morphemes that, at least in some of their uses, allow a “privative” meaning. This means that the compounds and phrases they form part of allow the proposition “(a) FAKE X is not (an) X” (Cappelle et al. 2018: 9). A *neppistool* ‘fake gun’, for instance, is not a type of gun, even though it shares certain properties with a true gun (Van der Wouden 2019). A *neppistool* may physically resemble a true gun, but may not be used to shoot metal bullets, for instance. Obviously, depending on the amount of shared properties with a “true” gun, some fake guns will be “faker” than others (think of a colorful plastic toy gun compared to a real-looking gun as used in movies). With respect to English *fake*, although often considered a textbook example of a privative adjective, Cappelle et al. (2018: 9) go so far as to contest “that fake is an across-the-board privative adjective, since a *fake article*, for instance, is most definitely still an article” (Cappelle et al. 2018: 9). A *fake article* is called *fake* because it lacks some of the prototypical properties of an authentic newspaper article, especially the property of providing true (i.e. fact-checked) content. Cappelle et al. (2018: 39) even show that *fake* is used in a privative sense in only half of the most frequent [*fake* + noun] combinations. Their study clearly demonstrates that the interpretation of *fake* is highly context-dependent. Not only the semantic properties of the noun play a crucial role in the interpretation of the combination of *fake* with a noun, but also the speakers’ cultural knowledge, as the following case illustrates:

To give another example, when there is discussion of *fake handbags*, what is meant is not things that look like handbags but that in fact lack the function of handbags; rather, *fake handbags* is more likely to be used in the sense of “real handbags that are passed off as being manufactured by a well-known designer brand”. This is based on our world knowledge about handbags, namely that they are not just functional objects but, perhaps in the first place, accessories whose make is important. (Cappelle et al. 2018: 40)

As stated above, we concentrate on a set of “fake” morphemes that are prone to privative use in contemporary Dutch, although they are not necessarily being used privatively across-the-board. We selected eight “fake” morphemes that are used in nominal or adjectival compounds, excluding privative adjectives (e.g. *vals* ‘false’, *schijnbaar* ‘apparent’, *kunstmatig* ‘artificial’) and affixes (e.g. *pseudo-*, *niet-* ‘not-’, *-achtig* ‘-ish, -like’). Below, we provide an example and the etymological source item¹ for each of these morphemes. *Nep* (1) originally comes from the German noun *Nepp* (‘unfair deal, fake copy’) or the German verb *neppen* ‘cheat’, but was already attested as a verb and noun in Dutch at the beginning of the 20th century. *Imitatie* (2) had been borrowed into Dutch from the French noun *imitation* around 1650. The items from (3) to (7) are indigenous Dutch morphemes. Lastly, we include the English loan *fake* since it can be used in Dutch compound-like sequences, as shown in (8).

- (1) ***nep-***: e.g. *nepjuwelen* ‘fake jewels’ (noun), *nepgouden* ‘fake gold’ (adjective)
 < *nep*_{V/N} ‘sham, fake, false’
- (2) ***imitatie-***: e.g. *imitatiekoffie* ‘surrogate coffee’ (noun), *imitatiefluwelen* ‘imitation velvet’ (adjective)
 < *imitatie*_N ‘imitation’
- (3) ***namaak-***: e.g. *namaakbont* ‘fake fur’

< *namaak*_V ‘imitate’

- (4) ***kunst***:- e.g. *kunstgras* ‘artificial grass’ (noun), *kunstleren* ‘imitation leather’
(adjective)

< *kunst*_N ‘art’

- (5) ***schijn***:- e.g. *schijnhuwelijk* ‘marriage of convenience’ (noun), *schijnheilig*
‘hypocritical’ (adjective)

< *schijn*_{V/N} ‘appear; appearance’

- (6) ***fop***:- e.g. *fopspeen* ‘dummy’, *fop-Jamaicaans* ‘fake Jamaican’ (adjective)

< *fop*_V ‘to fool sb.’

- (7) ***lok***:- *lokeend* ‘duck decoy’ (noun)

< *lok*_V ‘to lure sb.’

- (8) ***fake***:- e.g. *fake-bericht* ‘fake message’ (noun), *fakeblond* ‘fake blond’ (adjective)

< *fake*_{N/V/A}

As emerges from their etymologies, all these Dutch left-hand components originally derive from autonomous nouns and/or verbs. *Fake* was probably borrowed into Dutch as a noun, even though it already existed as an adjective and verb in English before that date.²

Interestingly, in synchronic corpus data several of these morphemes can be found in typically adjectival contexts as well (see Section 2). Since these adjectival uses are hardly mentioned in the dictionaries, we consider them as “innovative” – or even “extravagant” (see Section 2) – uses. For instance, in the *Van Dale* dictionary of Dutch *kunst*, *namaak* and *imitatie* are only listed as nouns, while *fop* and *lok* are only listed as verbs, and *schijn* occurs both as a verb and a noun. Only *fake* and *nep* have separate entries as adjectives. The following examples, all taken from the Dutch 2014 TenTen Web Corpus (*nlTenTen14*), illustrate the synchronic

orthographic and morphological variation of certain “fake” constructions (i.e. the combination *nep* ‘fake’ + *herinneringen* ‘memories’):

- (9) *Kwart mensen heeft **nepherinneringen***

‘Quarter of the people have **fake memories**’

- (10) *Plaatsen van **nep herinneringen** is mogelijk!*

‘Creating **fake memories** is possible!’

- (11) *Een onderzoek (...) stelde vast dat zelfs mensen met een "zeer ontwikkeld autobiografisch geheugen," (...), net zo makkelijk **neppe herinneringen** aanmaakten als iedereen.*

‘A study (...) found that even people with “highly developed autobiographical memory”, (...), created **fake memories** as easily as anyone.’

In (9) *nepherinneringen* forms a compound, and in (10) *nep* is orthographically separated from its head noun. Example (11) shows that *nep* can even be used as an inflected adjective. This synchronic variation is suggestive of a cline from compounding to adjectival use, in which the compound construction acts as a bridging context for the reanalysis of the nominal/verbal compound constituent into an adjectival modifier (cf. Van Goethem and De Smet 2014). This form of reanalysis has been identified before as an instance of “debonding” (Norde 2009), and is typically accompanied by category change (see further in Section 2).

The aim of this study is twofold. First, we compare the eight “fake” morphemes with respect to their semantic-distributional and morphological profiles to identify the specific functional and formal properties of each morpheme. Our second aim is to explore which factors trigger the “extravagant” behavior, i.e. the debonding and categorical flexibility, of some of these “fake” morphemes. As mentioned before, the English loan *fake* is included in this study

because it forms part of privative compound-like sequences in Dutch. However, since it was already attested as an adjective in English before being borrowed into Dutch (cf. footnote 2), we do not claim a parallel process of debonding and category change in the latter case, and include this morpheme mainly as a point of comparison for the semantic and distributional analysis of its Dutch counterparts.

On the methodological level, we will show how the combination of corpus analysis (extraction of comparable data samples, manual annotation of the semantic and morphological properties) with up-to-date statistical techniques allows us to detect both robust and more subtle differences between the “fake” morphemes. In addition, this mixed method enables us to identify the factors responsible for the recent morphological changes affecting some of them.

The results of our analysis will be theoretically embedded in the frame of Construction Grammar (cf. Hoffmann and Trousdale 2013). Crucial to this model is the concept of “constructions”, i.e. conventionalized form-meaning pairings, as the basic units of language. Constructions vary in size and complexity and exist at different levels of schematicity (i.e., levels of abstraction): a distinction can be made between fully schematic constructions (abstract grammatical patterns), semi-schematic constructions combining lexically filled positions with open slots, and substantive micro-constructions or fully idiomatic expressions. Based on Booij’s (2010) theory of Construction Morphology and the general principle in Construction Grammar that words are constructions (e.g. Goldberg 2006), we argue that our eight “fake” morphemes form part of semi-schematic morphological constructions with a filled and an open slot. The schema in (12) represents the compounding construction in which the “fake” morpheme is the left-hand member of a nominal or adjectival compound.³ As a rule, Dutch compounds are spelled as one word. However, in order to improve readability, a hyphen must be used in particular contexts, for instance to avoid a sequence of vowels at the constituent boundary (e.g. *imitatie-eiken* ‘imitation oak’) or in the case of special characters (capital letters,

numbers, symbols, etc.) on either side of the constituent boundary (e.g. *namaak-Vespa* ‘fake Vespa’).⁴ Nevertheless, actual language use contains many deviations from these rules, such that the spelling of compounds often seems to be a matter of free orthographic variation – compare *namaakeieren* ‘fake eggs’ in (12a) and *namaak-eieren* in (12b). In the schema in (13), on the other hand, the “fake” morpheme is written as a separate word, which may be considered another spelling variant, but which may also imply that the “fake” morpheme is used as an attributive adjective (13a), or as an adjective-modifying adverb (13b). Unambiguous examples of adjectivehood are inflected adjectives (such as (11) above), or cases with adverbial modification, such as *100%* as in (13a).⁵

(12) **compounding construction:**

$$[[FAKE](-)[a]_{N,A,i}]_{N,A,j} \quad \Leftrightarrow \quad [\text{fake SEM}_i]_j$$

a. *Liever dan hun eigen eieren, rollen meeuwen grote plastic **namaakeieren** hun nest in.*

‘Rather than their own eggs, seagulls are rolling big plastic **fake eggs** into their nests.’

b. *De pelikanen merken het verschil niet en gaan op de **namaak-eieren** zitten broeden.*

‘The pelicans don’t notice the difference and start breeding the **fake eggs**.’

(13) **modifying construction:**

$$[[FAKE][a]_{N,A,i}]_{NP,AP,j} \quad \Leftrightarrow \quad [\text{fake SEM}_i]_j$$

a. *Olli is de enige echte **100% namaak Olifant** ter wereld.*

‘Olli is the one and only **100% fake elephant** in the world.’

b. *We moeten dus niet **namaak blij** zijn [...].*

‘Hence we should not be **falsely happy**’

In Section 2, we will first discuss the concepts of “debonding” and “extravagance”. In Section 3, we present our case study of Dutch “fake” morphemes. The results of our analysis and our conclusions are presented in Sections 4 and 5 respectively.

2. “Fake” morphemes as “extravagant” morphology

2.1. Evaluative morphology

“Fake” morphemes are examples of evaluative morphology, which is a well-established field in morphological theory (e.g. Bauer 1997, Grandi and Körtvélyessy 2015). The concept of evaluation involves “a mental process by which objects of extra-linguistic reality are assessed from the point of view of quantity (big vs. small) and quality (good, bad, nasty, nice, nasty, etc.)” (Körtvélyessy 2014: 296). Thus far, the study of evaluative morphology has been largely restricted to the morphological expression of diminution and augmentation, and their pejorative or ameliorative connotations (e.g. Dressler and Barbaresi 1994), whereas other crucial functions such as the expression of “approximation” or “privative semantics” have been largely ignored. In the case of “fake” semantics, an object is lacking one or more of its prototypical features (see Section 1), which may result in a negative evaluation (bad quality). Therefore, it should also be considered an instance of evaluative semantics.

Evaluative morphology (partially) overlaps with “expressive morphology” (Zwicky and Pullum 1987, Szymanek 1988) and “extragrammatical morphology” (Dressler and Barbaresi 1994, Dressler and Karpf 1995, Mattiello 2013). While expressive morphology refers to word-formation with playful or poetic effects, the notion “extragrammatical” highlights that the morphological operations in this field often deviate from productive rules of word-formation

that allow a prediction of a regular output. Mattiello (2013), for instance, considers abbreviations (e.g. *BRB* = *Be Right Back*), blends (e.g. *beaulicious* < *beautiful* + *delicious*) and reduplicatives (e.g. *okey-dokey*) as instances of extragrammatical morphology, as these word-formation processes stand out as less predictable and therefore more creative and salient compared to derivation and compounding. This recalls Keller's (1994) maxim "talk in such a way that you are noticed", recast as the "maxim of extravagance" by Haspelmath (1999). In this vein, the term "extravagant morphology" has been coined to refer to any attention-attracting word-formation process (Eitelmann and Haumann 2019).

2.2. Debonding

Within the domain of evaluative morphology, it has already been shown that intensifying morphemes display "extravagant" morphological properties. Affixoids, i.e. morphemes "which look like parts of compounds, and do occur as lexemes, but have a specific and more restricted meaning when used as part of a compound" (Booij 2009: 208), are a case in point. In their study of three Germanic intensifying prefixoids (Dutch *kei* 'boulder', German *Hammer* 'hammer' and Swedish *kanon* 'cannon'), Norde and Van Goethem (2018) showed that these morphemes are subject to categorical flexibility to different degrees. For instance, German *Hammer* is not only used as an intensifying prefixoid (14), but may also act as an intensifying adverb (15) and as an (inflected) evaluative adjective (16) in colloquial usage (examples from COW14⁶, cf. Schäfer 2015):

(14) Intensifying prefixoid:

hammertrockene Haut 'extremely dry skin'

(15) Intensifying adverb:

*er hat es **hammer** gesungen* ‘he sung it **magnificently**’

(16) Evaluative (inflected) adjective:

*oh man ein total **hammeres** ende* ‘oh man, a totally **awesome** ending’

The preceding examples illustrate the process of “debonding”, which is “a composite change whereby bound morphemes (clitics, affixes, affixoids) in a specific context develop into free morphemes” (Norde 2009: 186). Several studies have demonstrated that debonding of intensifying prefixoids is a creative process of lexical innovation in Germanic languages, which may lead to the emergence of new intensifying adverbs or evaluative adjectives (cf. Norde and Van Goethem 2014, 2015, 2018; Van Goethem and Hiligsmann 2014, Van Goethem and De Smet 2014, Van Goethem and Hüning 2015; Battefeld et al. 2018). The present paper investigates to which extent the adjectival or adverbial uses of Dutch morphemes with “fake” semantics may be considered instances of debonding. We return to this issue in the next section, but first, let us have a look at the following examples, all drawn from the Dutch *nlTenTen14* web corpus (*Sketch Engine*, cf. Killgariff et al. 2014) (see also Section 3.1):

(17) *Eerlijk gezegd, ik serveer een **namaak vin santo**, want échte vin santo is ongelooflijk duur.*

‘To be honest, I serve a **fake vin santo**, because genuine vin santo is incredibly expensive.’

(18) *De uiteindelijke keuze is meestal een afweging tussen praktisch en mooi, echt of **imitatie**.*

‘The eventual choice is usually one between practical and nice, real or **imitation**.’

(19) *Meestal vind ik geuren in huidverzorgingsproducten **te nep** of chemisch ruiken*

‘Usually I find perfumes in skin care products smell **too fake** or chemical’

(20) *Dat is geen echte cupcake maar ik vond dat deze **neppe cupcake**, toch wel op de site mocht.*

‘That's not a real cupcake but I thought this **fake cupcake** was fine for the site.’

(21) *Bent u ook **nep aardig** of echt aardig?*

‘Are you also **fake nice** or really nice?’

(22) *hij begint te piepen en onnodig te **nep huilen***

‘he starts to squeak and unnecessarily to **fake-cry**’

(23) *Kim lijkt wel een steeds strakker en **faker gezicht** te krijgen.*

‘Kim looks like she is getting an ever tighter and **faker face**.’

In (17) *namaak* is found in attributive position, and in (18) *imitatie* is used in coordination with the adjective *echt* ‘real’. In example (19) *nep* is modified by the degree adverb *te* ‘too’, and in (20) the item features regular adjective inflection (*neppe*). Examples (21) and (22) show that *nep* can also be used as an adverbial modifier of an adjective or a verb. Finally, the comparative form of the loanword *fake* in example (23) indicates that *fake* may be used as an adjective in Dutch. All these examples show that certain “fake” morphemes may undergo debonding and categorical change to a more or lesser degree. The only possible exception is example (23), as the adjectival usage may have been borrowed from English directly (cf. footnote 2).⁷

2.3. Debonding as extravagant morphology

Since debonding involves the shift from a bound to a free morpheme, it is counterdirectional to grammaticalization, and can therefore be considered a case of degrammaticalization. Although

debonding is relatively rare where affixes and clitics are concerned (for examples see Norde 2009: 190–227), it appears to be more frequent in affixoid constructions (Norde and Van Goethem 2018). The privative morphemes discussed in the present paper are not affixoids (they do not have a meaning different from the free morpheme they derive from), but their use as compounding members is a prerequisite for the adjectival and adverbial constructions. As such, the novel usages of privative morphemes represent a type of debonding that has not been covered in previous studies.

In this paper, we argue that debonding (like degrammaticalization generally) is a type of extravagant morphology, which has two theoretical implications for the relation between extravagance and (de)grammaticalization. Firstly, as we will argue below, it is degrammaticalization, not grammaticalization, that is extravagant. Secondly, this means that extravagance cannot be invoked as an argument for the unidirectionality of change.

This challenges two major claims in a seminal paper by Haspelmath (1999), which introduces “the maxim of extravagance”. Referring to Keller (1994)’s maxim “talk in such a way that you are noticed”, Haspelmath (1999: 1043) argues that speakers introduce innovations in language to be noticed and, ultimately, to achieve social success. In Haspelmath’s view, grammaticalization is a “side effect of the maxim of extravagance, that is, speakers’ use of unusually explicit formulations in order to attract attention” (Haspelmath 1999: 1043).

Haspelmath introduces his maxim of extravagance in order to be able to explain why the two conflicting tendencies towards optimal clarity, on the one hand, and optimal economy, on the other “do not cancel each other out, leading to stasis rather than change” (Haspelmath 1999: 1052). In other words: if speakers want to get their message across in the most effective way, but at the lowest possible articulatory cost, why would languages change at all? After all, a more elaborate expression implies more effort, whereas a reduced expression is less effective.

Therefore, Haspelmath argues that speakers innovate because they want to be “socially successful”, following the maxim of extravagance. One example of such extravagant innovations is the use of motion verbs in future contexts, which eventually results in new periphrastic futures (such as *to be going to* in English or *aller* in French). Haspelmath goes on to argue that “degrammaticalization is by and large impossible because there is no counteracting maxim of “anti-extravagance” (Haspelmath 1999: 1043).

Although we agree with Haspelmath (1999) that there exists an asymmetry between grammaticalization and degrammaticalization in the sense that grammaticalization is more frequently attested and more cross-linguistically regular (Norde 2009), we do not think this asymmetry is due to the maxim of extravagance. Quite the reverse, we would argue that the pervasiveness and regularity of grammaticalization changes are not compatible with a conscious effort to be noticed (see Geurts 2000a, b for a careful argument why grammaticalization can be explained as the cumulative effect of communicative maxims and inter-individual variation, or Von Mengden and Kuhle (forthcoming), who convincingly argue that grammaticalization is an emergent phenomenon). Instead, we consider the changes involved in *degrammaticalization* as noticeable and extravagant, precisely because they run counter to the “predictable” path of grammaticalization. Instead of arguing that degrammaticalization is incompatible with extravagance, we claim, on the contrary, that it is particularly innovative and extravagant. Debonding of bound items and their categorical flexibility, in particular, are bound to be noticed.

3. Case study: “fake” morphemes in Dutch

3.1. Data and method

Our data were drawn from the Dutch 2014 TenTen Web Corpus (*nlTenTen14*), available on the *SketchEngine* platform (cf. Kilgariff et al. 2014).⁸ This corpus belongs to the TenTen corpus family, currently available in more than 30 languages, and consists of texts collected from the Internet. The *nlTenTen14* corpus contains 2,253,777,579 words. About three quarters of the Dutch corpus is extracted from the .nl domain; thus, the corpus mainly represents Netherlandic Dutch.⁹

We extracted random samples of 2000 tokens of each “fake” morpheme, making sure to include both its bound (compound) and free (debonded/adjectival) use. The CQL query allowed us to exclude a range of false positives a priori, such as derivatives beginning with the fake morpheme (e.g. *schijnbaar* ‘apparent’), homonymous verbs (e.g. *schijnen* ‘appear’) or homonymous nouns used in frequent collocations (e.g. *schijn bedriegt* ‘appearances are deceptive’, *geen schijn van kans* ‘no chance’). An example of the query we used to extract the *schijn*- constructions is given in (24):

(24) [word="schijn|schijn.+" & word!="schijnwerpers" & word!="schijnbaar" & word!="schijnbare" & word!="schijnen" & word!="schijnt"] [word!="bedriegt" & word!="je" & word!="op" & word!="te" & word!="tegen" & word!="van"]

Subsequently, we selected the first 500 relevant tokens pro morpheme, with the exception of *lok*, for which only 388 relevant instances were found. Each occurrence was manually annotated for its morphological and semantic properties in Excel. The statistical analyses were conducted in R (Levshina 2015, R Core Team 2016).

3.2. Semantic-distributional profile

When annotating the corpus examples, we noticed that quite a few occurrences contained sequences of different “fake” morphemes (25), and that some of them even appear to be interchangeable (26):

- (25) *laten we zwelgen in de plastic kerstbomen, de **kunstsneeuw**, **nep-champagne** en **namaaksterren** (...)*

‘let's wallow in the plastic Christmas trees, the **artificial snow**, **fake champagne** and **fake stars** (...)’

- (26) *Die wasabi moet **nep** zijn; hij is veel te goedkoop. Zelfs in Japan zou **namaak-wasabi** veel gebruikt worden.*

‘That wasabi must be **fake**; it's way too cheap. Even in Japan **fake wasabi** would be widely used.’

In order to compare the meaning and distribution of our eight “fake” morphemes, we make use of two different distributional approaches to semantics: Multiple distinctive collexeme analysis (3.2.1) and Semantic Vector Spaces (3.2.2).

3.2.1. Multiple distinctive collexeme analysis

In order to compare the meanings of our eight “fake” morphemes, we first performed a distributional-semantic analysis, namely Multiple Distinctive Collexeme analysis

(Stefanowitsch and Gries 2003; Levshina 2015: 248–251). This method is based on the co-occurrence frequencies of words that occur in near-synonymous constructions.

More specifically, the Multiple Distinctive Collexeme analysis compares the observed frequency of a specific slot filler (R1 in our case study, i.e. the item that the “fake” morpheme modifies: e.g. *wasabi* in [26]) to the expected frequency of that R1, and automatically computes if specific R1s are attracted to one of the constructions. We computed the Fisher exact test p -values to measure the attraction score between each morpheme and each R1. As a cut-off value for distinctiveness, we set 1.3 for log-transformed values (with 10 as the logarithm base), which approximately corresponds to the p -value of 0.05 (cf. Levshina 2015: 245).

Table 1 shows the top 20 of distinctive collexemes of *nep* (of a total of 46 distinctive collexemes). The most distinctive collexeme, *wimper* ‘eyelash’ occurs 19 times in combination with *nep*, only once in combination with *kunst*, and never with the remaining ‘fake’ morphemes. This makes it a significant distinctive collexeme of *nep* ($\log p > 15$). Although, as we will see in the next section, *nep* and *fake* form a cluster based on a Semantic Vector Spaces analysis and indeed share a number of collexemes (e.g. *profiel* ‘profile’, *tweet*, *site*), Table 1 also indicates that *nep* has a large number of distinctive collexemes with respect to *fake* (e.g. *wimper* ‘eyelash’, *nagel* ‘nail’, *brief* ‘letter’).

Table 1: Top 20 distinctive collexemes of *nep*

	R1	<i>fake</i>	<i>fop</i>	<i>imitatie</i>	<i>nep</i>	<i>kunst</i>	<i>schijn</i>	<i>lok</i>	<i>namaak</i>	$\log p$
1	<i>wimper</i> ‘eyelash’	0	0	0	19	1	0	0	0	15.807074
2	<i>nagel</i> ‘nail’	0	0	0	17	8	0	0	0	9.646586
3	<i>geld</i> ‘money’	1	0	1	9	0	0	0	2	5.398722
4	<i>brief</i> ‘letter’	0	0	0	5	0	0	0	0	4.460817
5	<i>wapen</i> ‘weapon’	0	0	0	5	0	0	0	0	4.460817
6	<i>factuur</i>	1	0	0	6	0	0	0	1	4.008330

	‘invoice’									
7	<i>borst</i> ‘breast’	0	0	0	4	0	0	0	0	3.567131
8	<i>beveiligings- software</i> ‘security software’	0	0	0	3	0	0	0	0	2.674208
9	<i>brood</i> ‘bread’	0	0	0	3	0	0	0	0	2.674208
10	<i>herinnering</i> ‘memory’	0	0	0	3	0	0	0	0	2.674208
11	<i>snor</i> ‘moustache’	0	0	0	3	0	0	0	0	2.674208
12	<i>tiet</i> ‘tit’	0	0	0	3	0	0	0	0	2.674208
13	<i>virusscanner</i> ‘virus scanner’	0	0	0	3	0	0	0	0	2.674208
14	<i>wijn</i> ‘wine’	0	0	0	3	0	0	0	0	2.674208
15	<i>ei</i> ‘egg’	0	0	0	5	0	0	0	4	2.554599
16	<i>haar</i> ‘hair’	2	0	0	4	0	0	0	0	2.484446
17	<i>medicijn</i> ‘medicine’	1	0	1	7	0	0	0	8	2.468868
18	<i>site</i> ‘site’	7	1	0	7	0	0	0	2	2.468868
19	<i>tweet</i> ‘tweet’	1	0	0	3	0	0	0	0	2.115966
20	<i>profiel</i> ‘profile’	30	0	0	12	0	0	0	4	1.888459

Table 2: Top 10 distinctive collexemes per “fake” morpheme

[INSERT TABLE 2 ABOUT HERE]

Table 2 shows the top 10 of distinctive collexemes (with a $\log p > 1.3$) for all “fake” morphemes. All in all, we found 56 distinctive collexemes for *fake*, 22 for *fop*, 42 for *imitatie*, 46 for *nep*, 16 for *kunst*, 46 for *schijn*, 22 for *lok* and 40 for *namaak*. However, these figures cannot be compared at face value, since they need to be checked against the type frequencies of each “fake” morpheme. We return to this issue in Section 3.2.3.

3.2.2. Semantic Vector Spaces

The Semantic Vector Spaces analysis (Levshina and Heylen 2014; Levshina 2015: 323–332) starts from the idea that words that occur in similar contexts tend to have comparable meanings, following Firth’s (1957) adage that “you shall know a word by the company it keeps” (Levshina and Heylen 2014: 22). The Semantic Vector Spaces (hence SVS) method is a radically data-driven technique that can be seen as an extension of collocational analysis, such as the one presented in the previous section (Levshina and Heylen 2014: 22). Semantic vectors are information strings that indicate “the weighted co-occurrence frequencies between target words and their contextual features” (Levshina 2015: 323). These frequencies are transformed into vectors, and the cosine of the angle between these vectors can be used as a measure of semantic proximity (Levshina and Heylen 2014: 23–24; Levshina 2015: 328–331). There are many variants of SVS analysis, depending on the number of target words and context size, as well as on the type of dependency relation between target word and context words (see Levshina and Heylen 2014: 25–29 for details). Our study is a small-scale type of SVS with only eight target words and a minimal context (R1). The advantage of small contexts in SVS analysis is that they tend to result in tighter taxonomic relations (Levshina and Heylen 2014: 26). Applied to our case study, SVS analysis is used to identify clusters of (near-)synonyms, based on the R1s of each “fake” morpheme.

Table 3 is a similarity matrix that indicates the degree of similarity between the semantic vectors of each “fake” morpheme. The cosine values were computed using the `cossim` function in the `Rling` package (Levshina 2014). The resulting scores range from 0 (no overlap at all between the R1s) and 1 (complete overlap of the R1s). Obviously, every morpheme has a perfect collocational overlap with itself. This is shown by the diagonal scores in the matrix that are all equal to 1. The matrix (Table 3) furthermore reveals some interesting patterns:

- *fake/nep* have the highest similarity scores (> 0.06);

- *nep/namaak* (> 0.05) and *imitatie/namaak* (> 0.04) are relatively similar to each other;
- the morphemes *fop*, *kunst*, *lok* and *schijn* display low cosine similarity towards the other morphemes (< 0.03);
- no collocational overlap at all is found between the pairs *fop/kunst*, *fop/lok*, *imitatie/lok* and *kunst/fop*.

Table 3: Cosine similarity between the “fake” morphemes based on R1

	fake	fop	imitatie	nep	kunst	schijn	lok	namaak
fake	1.0000	0.0127	0.0060	0.0604	0.0015	0.0212	0.0116	0.0234
fop	0.0127	1.0000	0.0034	0.0292	0.0000	0.0142	0.0000	0.0192
imitatie	0.0060	0.0034	1.0000	0.0325	0.0196	0.0153	0.0000	0.0455
nep	0.0604	0.0292	0.0325	1.0000	0.0142	0.0096	0.0049	0.0525
kunst	0.0015	0.0000	0.0196	0.0142	1.0000	0.0077	0.0055	0.0115
schijn	0.0212	0.0142	0.0153	0.0096	0.0077	1.0000	0.0044	0.0064
lok	0.0116	0.0000	0.0000	0.0049	0.0055	0.0044	1.0000	0.0050
namaak	0.0234	0.0192	0.0455	0.0525	0.0115	0.0064	0.0050	1.0000

Once these cosine similarity measures have been established, they can be transformed into distances, using the `pam` function in the `cluster` package (Maechler et al. 2019; for details on this method see Levshina 2015: 330–331). This method enables us to draw a cluster dendrogram (Figure 1) that provides a visualization of the results of the SVS analysis. In order to obtain the optimal number of clusters, this function computes the “average silhouette width”, which ranges from 0 (no clusters at all) to 1 (perfect separation of all clusters; Levshina 2015: 311). The highest average silhouette width is obtained with 7 clusters, although it is still quite low (0.25). The resulting cluster dendrogram in Figure 1 is drawn using the `ward.D2` algorithm, which minimizes the increase in the variance in the distances between the members of the

clusters (Levshina 2015: 311). As Figure 1 shows, *fake* and *nep* cluster together, whereas *imitatie* and *namaak* belong to the same overarching branch. The other 4 morphemes, *kunst*, *lok*, *fop* and *schijn*, all form individual branches, which indicates that they have unique semantic-distributional properties.

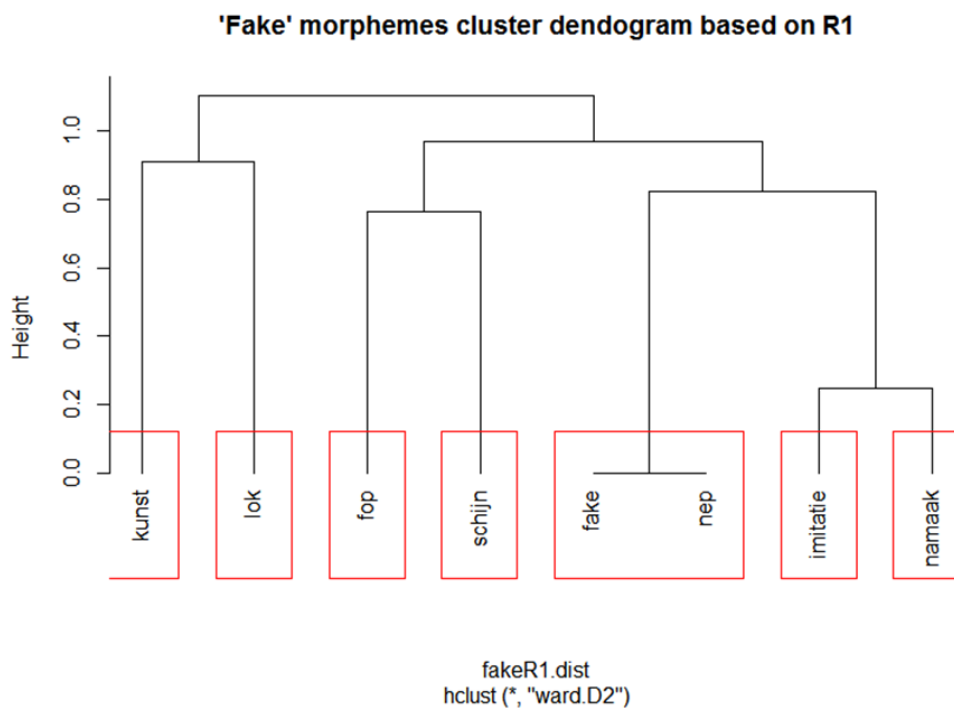


Figure 1: “Fake” morphemes cluster dendrogram based on R1

3.2.3. Semantic-distributional profile: discussion

The semantic-distributional analyses explained in the preceding sections have first of all shown that *fake* and *nep* are most similar semantically: their semantic-distributional profiles are most alike, albeit with still a relatively high number of specific collexemes. These morphemes stand out as the most “neutral” “fake” morphemes, and *nep* could be considered the Dutch equivalent of English *fake*. A closer look at the dataset shows that *fake* and *nep* are indeed quite often

interchangeable (27–28), but that – unsurprisingly – *fake* more often collocates with English loanwords than *nep* (e.g. *fake fur*, *fake account*, *fake speech*).¹⁰

(27) *Een echte of een **fake kerstboom**? Ik heb samen met mijn vriend een **nep kerstboom** gekocht.*

‘A real or a **fake Christmas tree**? Together with my friend I bought an **artificial Christmas tree**.’

(28) *Wat zijn Tijdelijke Tattoos? Veel mensen noemen ze **fake tattoos**, plaktattoos of **neptattoos**.*

‘What are Temporary Tattoos? Many people call them **fake tattoos**, stick tattoos or **artificial tattoos**.’

A second conclusion is that *imitatie* and *namaak* also have relatively comparable distributional-semantic profiles in terms of semantic vectors. A comparison of their distinctive collexemes (see Table 2), however, reveals that *imitatie* primarily refers to faithful copies of expensive natural materials (leather, fur, caviar, marble, etc.), whereas *namaak* more often has a connotation of fraud (medicine, article, cigarettes, or brand names such as *Uggs*).

Finally, the remaining four “fake” morphemes have more specific and unique semantic-distributional profiles. Based on their most distinctive collexemes (Table 2), their constructional meanings can be described as follows:

- *kunst*-constructions refer to an “artificial, synthetic, not natural X” (e.g. *kunststof* ‘plastic’, *kunstgras* ‘artificial turf’);
- *schijn*-constructions refer to an “apparent X”; *schijn*- expresses a superficial resemblance (e.g. *schijnzwangerschap* ‘phantom pregnancy’), often accompanied

by a pejorative connotation (e.g. *schijnheilig* ‘hypocritical’, *schijnveiligheid* ‘false safety’);

- *fop*-constructions refer to a “fake X with the intention to fool somebody, meant as a joke” (e.g. *fopspeen* ‘dummy’¹¹, *fopneus* ‘fake party nose’, *fopgesprek* ‘joke interview’);
- finally, *lok*-constructions refer to a “fake X in order to lure somebody, intended as a trap” (e.g. *lokvogel* ‘decoy (bird)’, *lokfiets* ‘bait bike’).

These conclusions are corroborated by the distinctive collexeme ratios of each morpheme, i.e. its number of distinctive collexemes divided by its number of types in the corpus sample. Figures 2 and 3 show that *fake*, *imitatie*, *nep* and *namaak* have the lowest distinctive collexeme ratios, while *fop*, *kunst*, *lok* and *schijn* have the highest scores. This means that the former morphemes have relatively few distinctive R1s with respect to their (high) number of types, whereas the latter morphemes have a relatively high proportion of distinctive R1s on their total (low) number of types. The former morphemes can therefore be considered to have more “neutral” “fake” semantics, while the latter four have more specific semantic-distributional profiles. In the case of *kunst* and *lok*, for instance, more than half of their number of types are distinctive with respect to the other “fake” morphemes.

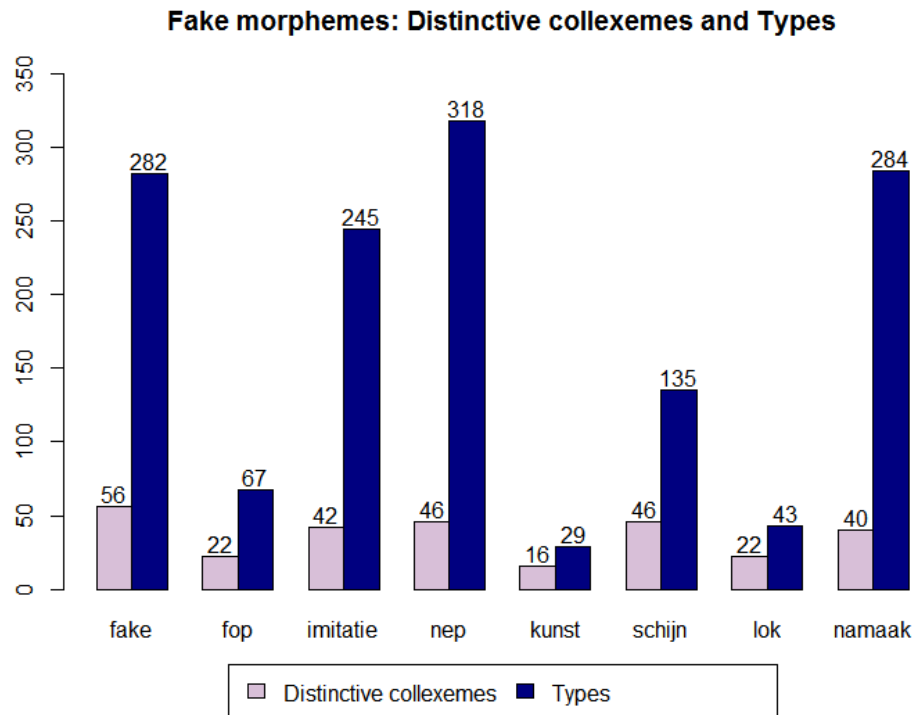


Figure 2: Number of distinctive colllexemes and number of types per “fake” morpheme

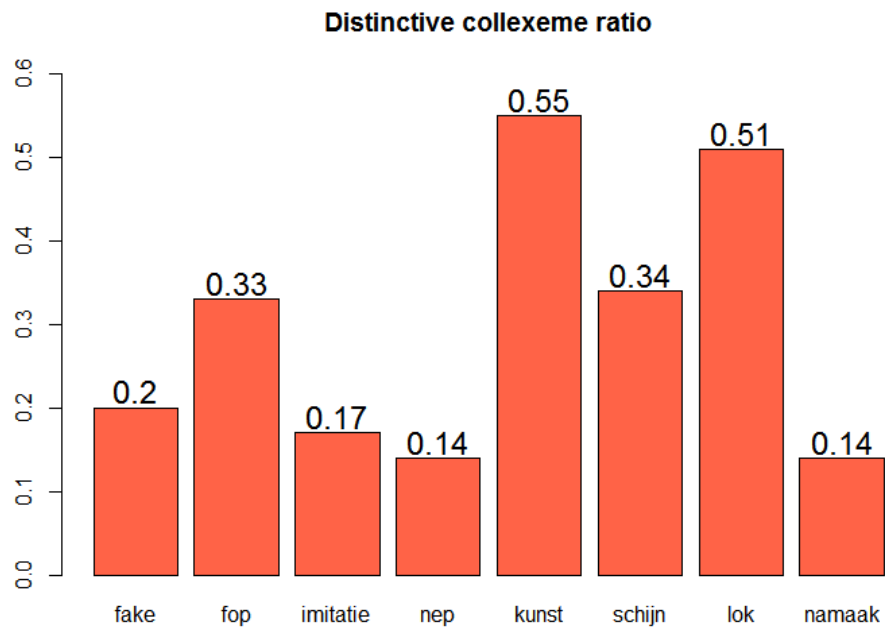


Figure 3: Distinctive colllexeme ratio per “fake” morpheme

3.3. Debonding and categorical flexibility

Having examined the semantic-distributional behavior of the eight “fake” morphemes, we now turn to their morphological properties. More specifically, we are interested in the extent to which each morpheme allows debonding and categorical flexibility. We consider debonding here as a purely orthographic criterion; it implies that the compound is written apart (e.g. in *schijn democratie* ‘fake democracy’). Sequences written as one word (e.g. *kunstgebit* ‘false teeth’) and hyphenated sequences (e.g. *imitatie-Vespa* ‘counterfeit Vespa’) have been considered instances of bound use (see Section 1 on the use of hyphenation in Dutch compounds). In our comparison, we include the English loan *fake*, even though it may be argued that sequences such as *fake tattoo* were borrowed directly and hence have never been written together.¹² This implies that, strictly speaking, we cannot be certain whether free *fake* is indeed an instance of debonding, but it may nevertheless be interesting to compare the spelling variation of *fake* constructions to the other seven. The *fake* frequencies are therefore included in the tables and figures below, with the proviso that separate *fake* does not necessarily reflect debonding.

Categorical flexibility, on the other hand, refers to all morphological or syntactic features that flag “adjectivehood”: predicative use, gradation, inflection, coordination with another adjective, scope over an NP, verb or adjective modification (in the case of adverbial use) (cf. examples 17–23).

Table 4 and Figure 4 indicate the ratios for each morpheme.

	Debonding ratio	Adjectivehood ratio
<i>fake</i>	372/500 = 0.744	77/500 = 0.154

<i>fop</i>	15/500 = 0.030	0
<i>imitatie</i>	179/500 = 0.394	40/500 = 0.080
<i>kunst</i>	0	0
<i>lok</i>	3/388 = 0.008	0
<i>namaak</i>	152/500 = 0.304	44/500 = 0.088
<i>nep</i>	150/500 = 0.300	93/500 = 0.186
<i>schijn</i>	15/500 = 0.020	2/500 = 0.004

Table 4: Debonding and adjectivehood ratios per “fake” morpheme

Fake morphemes: Adjectivehood and Debonding ratios

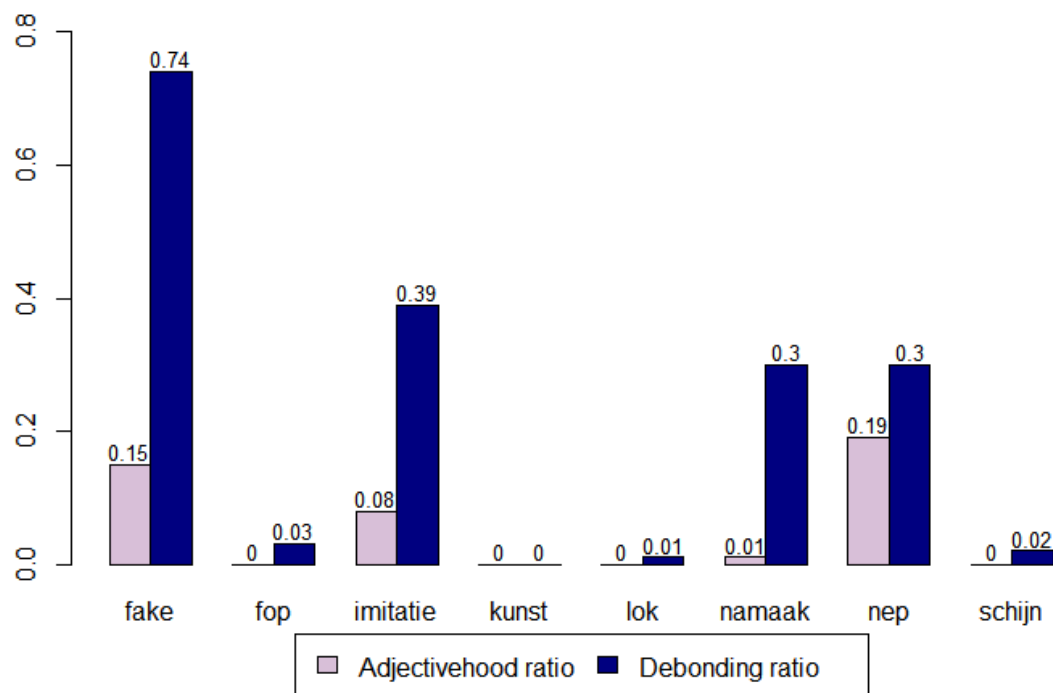


Figure 4: Debonding and adjectivehood ratios per “fake” morpheme

Strikingly, we observe a similar dichotomy between two groups of morphemes as we found in the semantic-distributional analysis: *fake*, *imitatie*, *nep* and *namaak* have relatively high degrees of debonding and/or adjectivehood, whereas *fop*, *schijn*, *lok* and *kunst* do not show this “extravagant” morphological behavior. The latter morphemes are (almost) exclusively found in their bound forms (e.g. *fopwetenschapper* ‘pseudo-scientist’, *schijnbeweging* ‘feint (movement)’, *lokagent* ‘fake policeman’, *kunstgebit* ‘false teeth’) and do not show any signs of adjectivehood.

Imitatie and *namaak* have a relatively high degree of debonding (> 30%). This may be due to their length and the fact that they form relatively complex compounds (e.g. *imitatie sneeuwkettingen* ‘imitation snow chains’, *namaak merkartikelen* ‘counterfeit branded goods’). When they combine with (capitalized) proper names referring to brands, debonding also occurs: for instance, *namaak Gucci* ‘counterfeit Gucci’, *imitatie iPad* ‘imitation iPad’. However, both morphemes show few signs of adjectivehood: mostly some predicative use (29), but no inflection at all.

(29) *Taarten met toppings, zo mooi dat het wel **imitatie** lijkt.*

‘Cakes with toppings, so beautiful it resembles **imitation**.’

Applying the spelling criterion to *fake*, this morpheme is written separately most often (74.4%), but, as we argued above, this is due to the fact that it is an English loanword and is not bound in English “fake” constructions, such as *fake account*. When the R1 is a Dutch word, it is bound in about a quarter of the occurrences in the sample (e.g. *fake-advertentie* ‘fake advertisement’). Its degree of adjectivehood in Dutch, mostly manifested through predicative use, as in (30), is however rather low (15.4%).

(30) *Ik wist direct dat de mail **fake** was omdat deze toekwam op een mailadres dat niet gelinkt is aan mijn Facebookprofile.*

‘I knew immediately that the mail was **fake** because it arrived at an email address that is not linked to my Facebook profile.’

An important proviso is in order at this point. Since *fake* ends in *-e*, which is the ending of inflected Dutch adjectives, we have been unable to determine the presence of inflection of *fake* in our written data. In spoken Dutch however, *fake* may be inflected (in which case it is pronounced as two syllables, not one), as shown in Goublomme’s (2019) study of the integration of *fake* in both written and spoken Dutch. Her analysis of survey data from 185 native speakers revealed that almost half of them inflect *fake* in spoken discourse in syntactic contexts where Dutch adjective inflection is required, as in (31).¹³ Based on these facts, we may assume that the degree of adjectivehood of *fake* in our corpus sample is in reality somewhat higher than the reported ratio.

(31) *Ze heeft veel vrienden, maar ook veel **fake vrienden**.*

‘She has many friends, but also many **fake friends**.’

Finally, *nep* stands out as the Dutch “fake” morpheme with the highest degree of adjectivehood (18.6%). More than 60% of its debonded uses involve signs of categorical flexibility, including inflection and gradation, as in (32–33).

(32) *Maar blijkbaar gelden deze normen en waarden niet als je de tijd van de dinosauriërs nog hebt meegemaakt, je gebit nog **nepper** is dan de borsten van Kim Kardashian (...)*

‘But apparently these norms and values don't apply if you have experienced the time of the dinosaurs, if your teeth are even **faker** than Kim Kardashian’s breasts (...)’

(33) *Maar als je echt een mooi boek wilt lezen, raad ik hem af. Daarvoor is hij iets **te nep**, de gebeurtenissen zijn te doordacht.*

‘But if you really want to read a beautiful book, I do not recommend it. It’s a bit **too fake** for that, the story is too much thought-out.’

To sum up, so far we noticed that the “fake” morphemes with the most neutral “fake” semantics undergo debonding more often and show more categorical flexibility than the ones with more specific semantic-distributional profiles. In the next section, we analyze if the same dichotomy applies to the productivity of the morphemes under study.

3.4. Productivity

In order to establish whether there exists a correlation between debonding and productivity, we computed both the type / token ratio (TTR) and the potential productivity (PP)¹⁴ of all eight “fake” morphemes.

Morph	TTR	PP
<i>fake</i>	282/500 = 0.564	218/500 = 0.436
<i>fop</i>	67/500 = 0.134	43/500 = 0.086
<i>imitatie</i>	245/500 = 0.490	192/500 = 0.384
<i>kunst</i>	29/500 = 0.058	10/500 = 0.02

<i>lok</i>	43/388 = 0.11	21/388 = 0.05
<i>namaak</i>	284/500 = 0.568	223/500 = 0.446
<i>nep</i>	318/500 = 0.636	249/500 = 0.498
<i>schijn</i>	135/500 = 0.27	88/500=0.18

Table 5: Productivity of the “fake” morphemes

The results are shown in Figure 5.

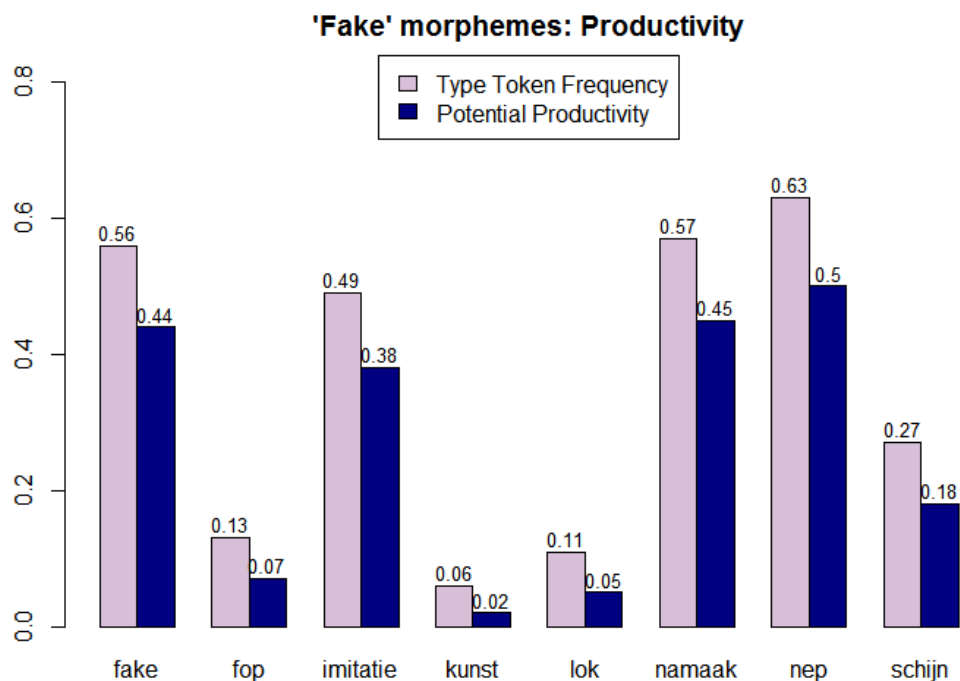


Figure 5: Productivity of the “fake” morphemes

The productivity ratios tally with the figures for distinctive collexemes discussed in Section 3.2. On the one hand, *nep*, *fake*, *namaak* and *imitatie* show relatively high productivity ratios, both regarding the number of different types they create (TTR) and the number of hapax legomena (PP) (e.g. *nepzwanger* ‘falsely pregnant’, *namaakwielrenner* ‘fake (amateurish) cyclist’).

Schijn, *fop*, *lok* and *kunst*, on the other hand, are the least productive morphemes. *Nep* is the most productive “fake” morpheme and *kunst* the least productive one. Illustrative of this difference in productivity is the fact that *nep*- forms 318 different types among which 249 are hapaxes in the data set of 500 occurrences (TTR = 0.63, PP = 0.50), while *kunst*- features only 29 different types, among which 10 are hapaxes (TTR = 0.06, PP = 0.02). Moreover, the most frequent type *kunststof* ‘synthetic material’ covers 64.4% (322/500) of its tokens. These differences in frequency distributions are illustrated in Figure 6.

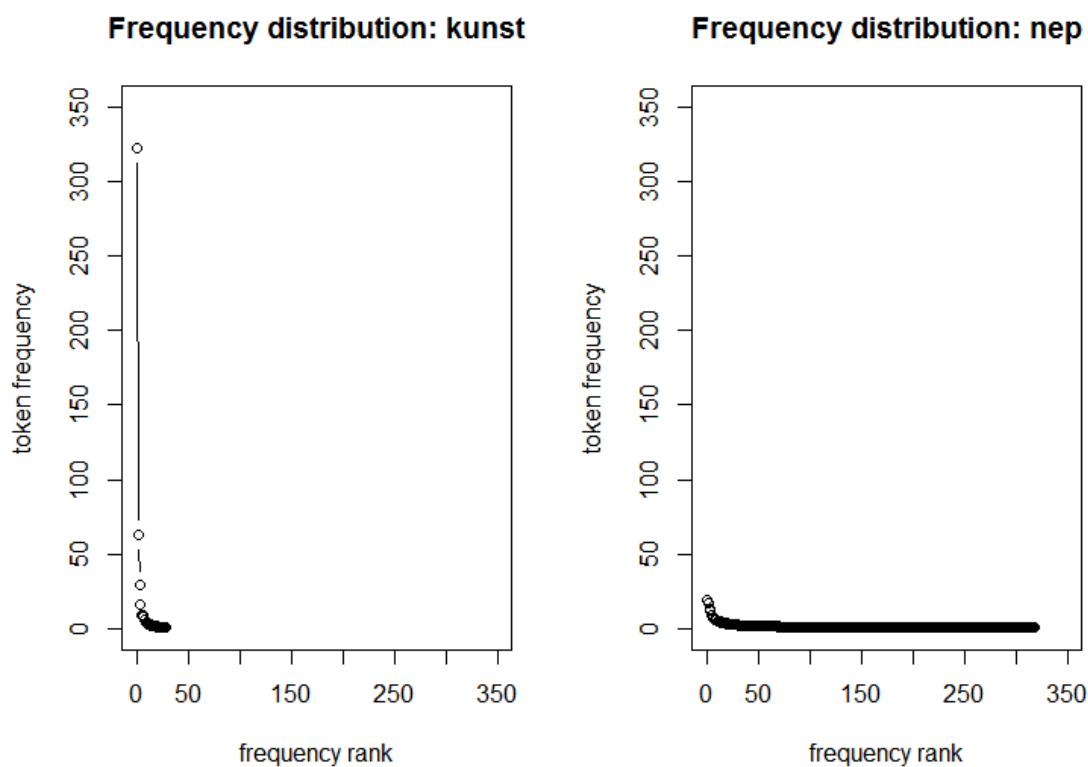


Figure 6: Frequency distributions: *kunst* vs *nep*

Given these differences, it can be useful to plot potential productivity against type frequency, a method proposed in Baayen and Lieber (1991: 818–819), which they term “global

productivity”. Global productivity of the eight “fake” morphemes is shown in Figure 7, with potential productivity on the y-axis and type frequency on the x-axis. This figure shows a clear cline from *nep*, which has highest global productivity, to *kunst*, which has lowest global productivity.

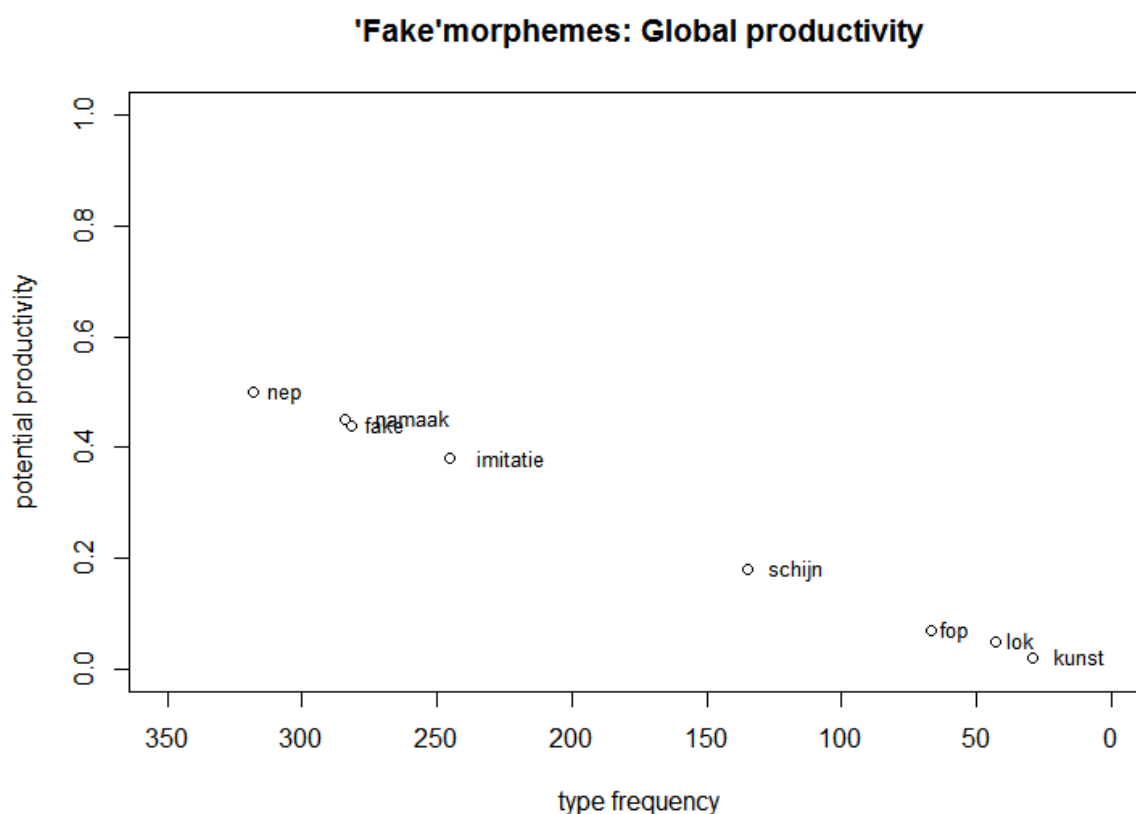


Figure 7: Global productivity of the “fake” morphemes

4. Discussion of the results

From the preceding analyses, it is clear that two groups of “fake” morphemes can be distinguished based on their semantic-distributional profile, their morphological properties and their productivity. *Fake*, *nep*, *namaak* and *imitatie* are semantically close to each other, allow a relatively high degree of debonding and – albeit to a lesser extent – of categorical flexibility.

Moreover, they are productive morphemes. *Schijn*, *fop*, *lok* and *kunst* have semantically distinctive profiles, hardly allow debonding or adjectival use and show low productivity. In the next three sections, we will explore the relation between these three parameters in a stepwise fashion. Section 4.1 focuses on the relation between productivity and debonding, Section 4.2 studies how the semantic profile relates to productivity, and Section 4.3 investigates the relationship between the three parameters.

4.1. Productivity and debonding

In order to answer the question whether increasing productivity significantly correlates with increasing debonding, we performed a correlation analysis with linear regression modelling (Levshina 2015: 115–170) between TTR and debonding ratio. Since debonding of *fake* was already predicted by its English origin, this item was excluded from the analysis.

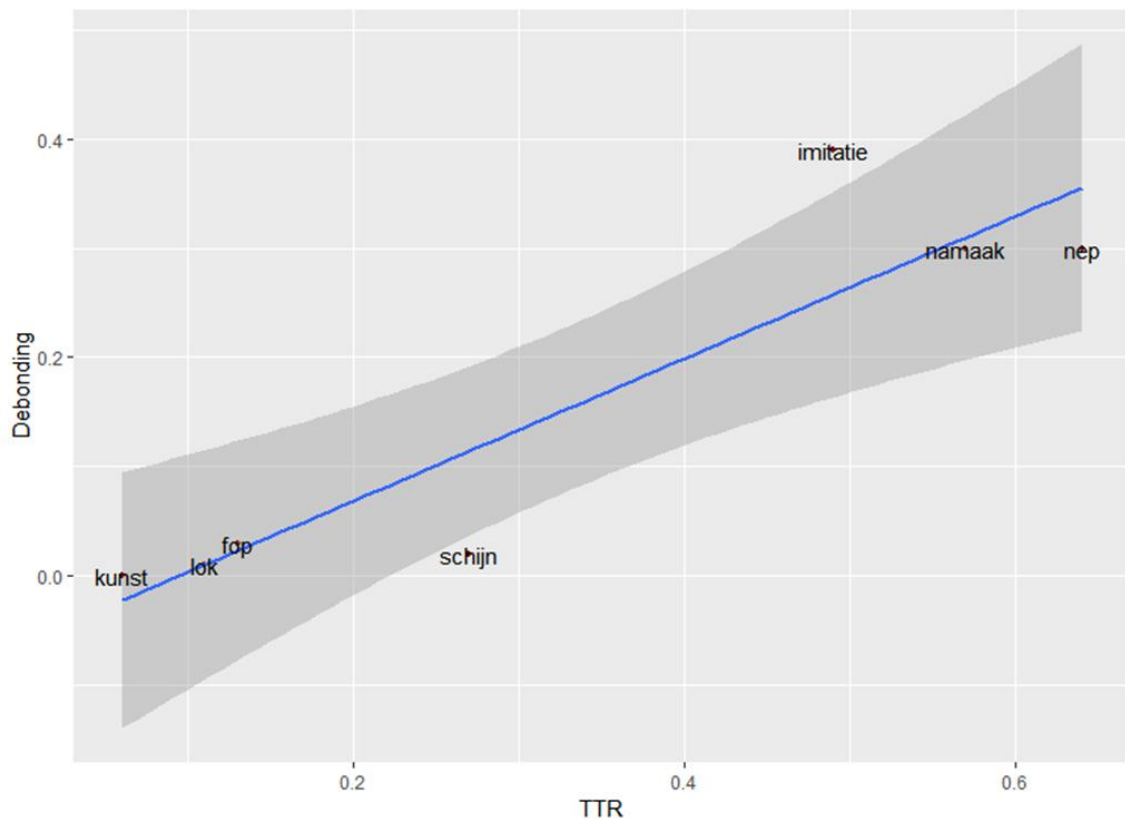


Figure 8: Correlation between TTR (explanatory variable) and debonding ratio (response variable)

Figure 8 indicates a significant correlation between productivity (TTR) and the extent of debonding of the seven morphemes ($p = 0.004396 (**)$; Adjusted $R^2 = 0.8288$; cor (Pearson's) = 0.910). This result implies that increase in productivity is accompanied by increase in debonded use. Again, *kunst* on the one hand and *nep* on the other can be identified as the two opposite endpoints of the cline.

4.2. Semantic coherence and productivity

In order to explain this positive correlation between productivity and debonding, we need to take into account a third parameter, namely semantic coherence. Barðdal (2008) claims that there exists an inverse correlation between type frequency and semantic coherence of a construction, as represented in Figure 9:

[...] the higher the type frequency of a construction, the lower the degree of semantic coherence is needed for a construction to be productive. Conversely, the lower the type frequency of a construction, the higher degree of semantic coherence is needed for a construction to be extendable. (Barðdal 2008: 35)

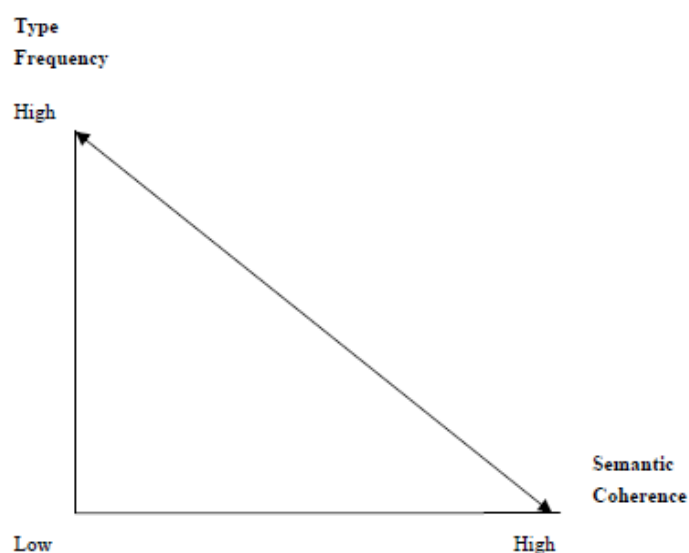


Figure 9: Inverse correlation between semantic coherence and type frequency
(Barðdal 2008: 35)

Partially filled constructions (or constructional idioms) typically impose formal and/or semantic restrictions on the items that can fill their open slots, and these restrictions constrain the productivity of the construction (see, for instance, Gyselinck and Coleman (2016) on the pseudo-reflexive resultative construction in Dutch).

We argue that this inverse correlation between type frequency and semantic coherence also holds for our “fake” morphemes. *Nep*, *fake*, *namaak* and *imitatie* have the highest type frequency. Additionally, the semantic-distributional analysis in Section 3.2 indicated that *nep* and *fake* are semantically the most neutral “fake” morphemes, followed by *imitatie* and *namaak* which display some privileged connotations. These four morphemes show the lowest distinctive collexeme ratios, which we consider a proxy for their (relatively low) semantic coherence. *Schijn*, *fop*, *lok* and *kunst* have been shown to display low type frequency and high distinctive collexeme ratio, which suggest a relatively high semantic coherence: in Section 3.2.3 we have identified the specific semantic restrictions each of these four morphemes imposes on its R1. Consequently, the eight “fake” morphemes can be ranged on Barðdal’s cline according to their degree of semantic coherence and type frequency. However, Barðdal’s correlation in Figure 9 is schematic, so it might be interesting to explore whether an actual regression analysis yields a similar picture. Therefore, to visualize the relation between type frequency and semantic coherence based on our corpus data, we performed a second regression analysis, in which we took the distinctive collexeme ratio (see Figure 3 in Section 3.2.3) as a measure for semantic coherence. The regression line in Figure 10 bears a strong resemblance to Barðdal’s schematization, reflecting a strong inverse relationship (the correlation is negative): $p = 0.0004905$ (***); Adjusted $R^2 = 0.8664$; $\text{cor (Pearson’s)} = -0.9410214$.

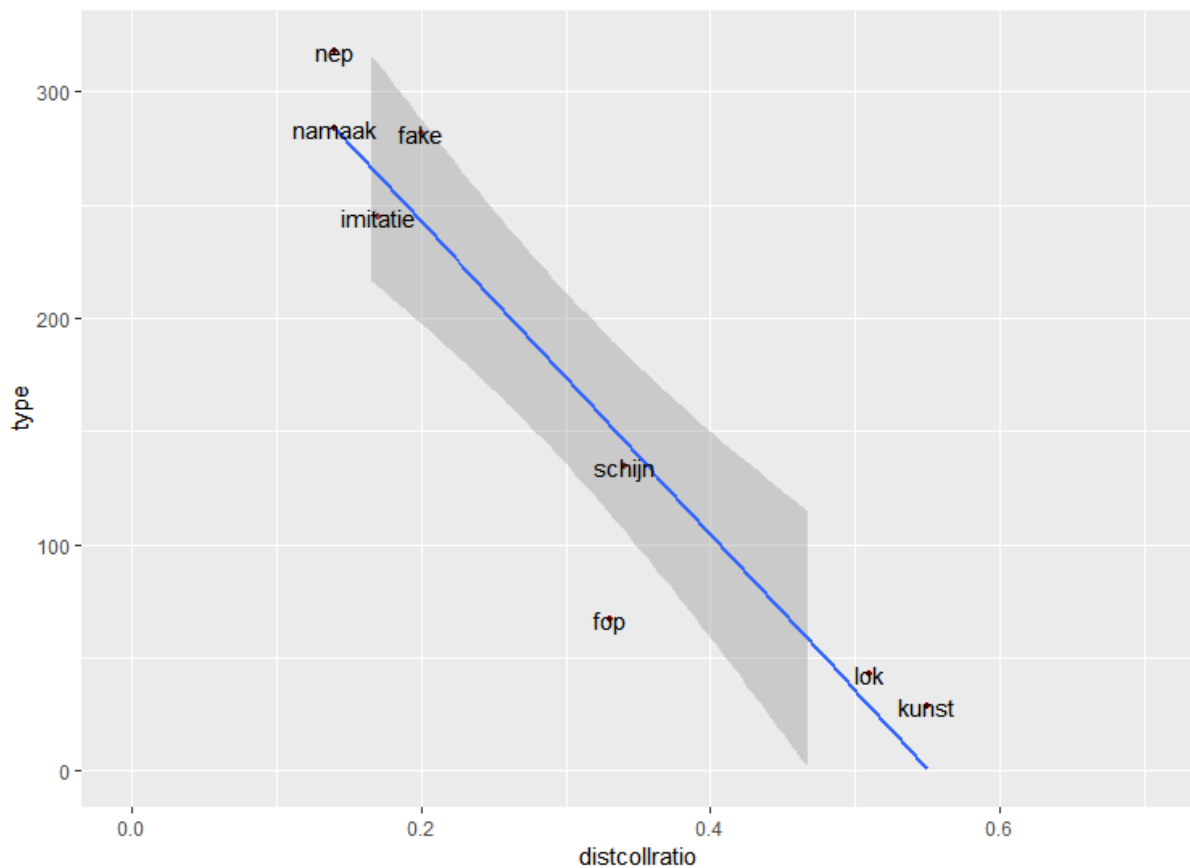


Figure 10: Inverse correlation between distinctive collexeme ratio (explanatory variable) and type frequency (response variable)

4.3. Semantic coherence, productivity and debonding

In the preceding sections, we have shown that the type frequency of the “fake” morphemes positively correlated with their degree of debonding and that there exists an inverse correlation between their distinctive collexeme ratio and type frequency, the former taken as a proxy of semantic coherence. This implies that high type frequency goes together with low semantic coherence and high degree of debonding, and, conversely, that low type frequency correlates with high semantic coherence and low degree of debonding.

Earlier research suggests an inverse correlation between semantic coherence and debonding. For instance, Battfeld et al. (2018) show that evaluative prefixoids such as German *Spitze* ‘top’, Dutch *top* ‘top’ or Swedish *skit* ‘shit’ debond far more often than compounding elements that have a specific meaning. Similarly, Norde and Van Goethem (2018: 489) found that evaluative prefixoids have a far higher debonding ratio than their equivalents in simile constructions. For example, Dutch *kei* (literally ‘boulder’) is separated from the adjective with which it collocates in 48% of all compounds where *kei* has intensifying meaning (e.g. *kei vatbaar* ‘very prone’), but only in 12% of collocations where it has simile meaning (e.g. *keihard* ‘boulder hard’, i.e. ‘hard as a boulder’). In both these studies, the evaluative / intensifying prefixoids have lower semantic coherence due to substantial semantic bleaching, which goes hand in hand with higher debonding ratios. We assume that the same holds true for the Dutch “fake” morphemes. Low semantic coherence, as in the case of the constructions with *nep*, *fake*, *namaak* and *imitatie*, may result in weak morphological coherence and trigger debonding, potentially followed by adjectival reanalysis. Conversely, high semantic coherence, as attested in the constructions with *schijn*, *fop*, *lok* and *kunst*, may correlate with strong morphological coherence and prevent debonding of the “fake” morpheme.

Adopting a constructionist point of view (cf. Section 1), we claim that the eight Dutch “fake” constructions can be ranged on a cline from [+schematic] to [+substantive]. The *nep*-construction, for instance, shows low semantic coherence and high type frequency, and can therefore be considered more “schematic” than, for instance, the *kunst*-construction. The latter forms part of semantically coherent compounds with a limited number of types and can therefore be considered more “substantive”. Interestingly, only the most schematic constructions of the cline undergo debonding: they form productive schemas with few semantic restrictions on their open slot and therefore resemble syntactic patterns of the attributive [ADJ N]_{NP} type (e.g. *neppasppoort/ vals paspoort* ‘fake/false passport’). This similarity may trigger

the adjectival reanalysis of a morpheme such as *nep*. The least schematic constructions of the cline occur as coherent, often lexicalized compounds (e.g. *fopspeen* ‘dummy’, *kunststof* ‘synthetic’): they clearly belong to the domain of lexical morphology – and not of syntax – and hence do not undergo debonding or reanalysis into a syntactic construction.

Figure 11 summarizes this relation between semantic coherence (distinctive collexeme ratio), type frequency and debonding.

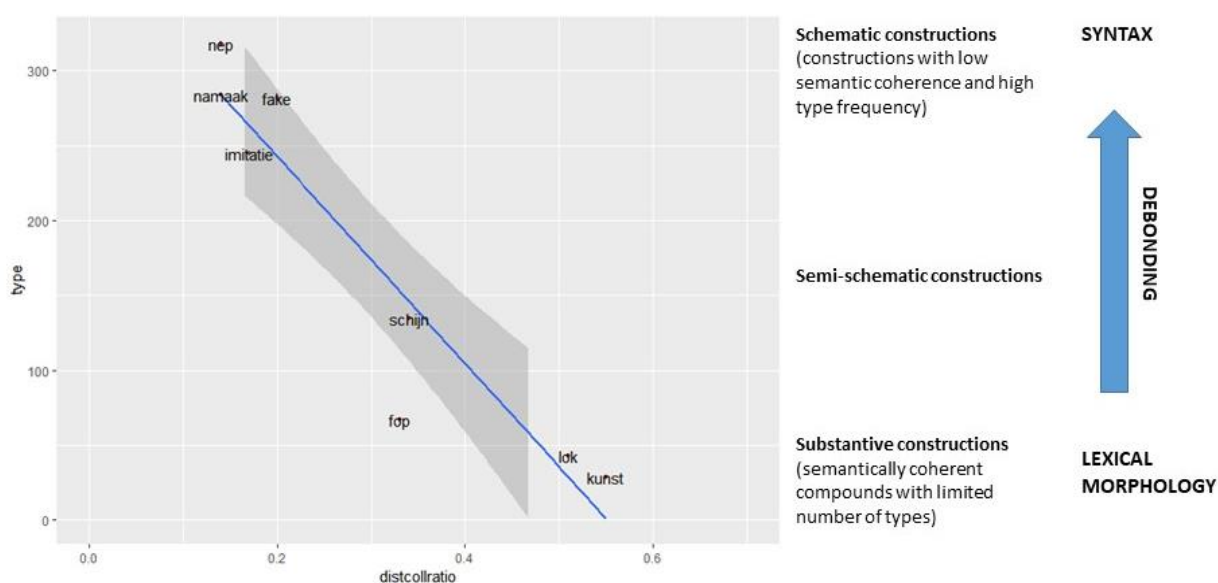


Figure 11: Semantic coherence, type frequency and debonding of the eight “fake” morphemes

5. Conclusions

In this paper, we have concentrated on the “extravagant” properties of evaluative morphology. More specifically, we have argued that debonding and categorical flexibility of a set of Dutch “fake” morphemes should be considered an instance of extravagant morphology. Consequently,

we refute Haspelmath's (1999) claims that extravagance is a property of grammaticalization and that, as a result, degrammaticalization cannot occur.

Through a combination of an extensive corpus analysis and statistical methods (Semantic Vector Spaces, Multiple Distinctive Collexeme analysis, Regression), we have explored the factors that trigger the debonding and categorical flexibility of some Dutch “fake” morphemes, and that prevent others from debonding.

Our analyses show that this “extravagant” behavior is triggered by an interplay of two factors: type frequency and semantic coherence. These parameters allow us to range the eight “fake” morphemes on a cline from [high type frequency] / [low semantic coherence] to [low type frequency] / [high semantic coherence]. We argue that morphemes belonging to the upper part of the cline (*nep*, *fake*, *namaak*, *imitatie*) undergo debonding more easily due to their low degree of semantic and morphological coherence and their resemblance to the syntactic/schematic attributive construction (cf. the cline *nepherinneringen* / *nep herinneringen* / *neppe herinneringen* / *valse herinneringen* ‘fake/false memories’). Conversely, morphemes belonging to the lower part of the cline display strong semantic and morphological coherence and the (lexicalized) compounds they form are not subject to debonding or further reanalysis into adjectives.

References

- Baayen, Harald & Rochelle Lieber. 1991. Productivity and English derivation: A corpus-based study. *Linguistics* 29. 801–843.
- Baayen, Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics. An international handbook*, 900–919. Berlin & New York: Mouton de Gruyter.
- Barðdal, Jóhanna. 2008. *Productivity: Evidence from case and argument structure in Icelandic*. Amsterdam & Philadelphia: John Benjamins.
- Battefeld, Malte, Torsten Leuschner & Gudrun Rawoens. 2018. “Evaluative Morphology” in German, Dutch and Swedish: Constructional Networks and the Loci of Change. In Kristel Van Goethem, Muriel Norde, Evie Coussé & Gudrun Vanderbauwhede (eds.), *Category change from a constructional perspective*, 229–262. Amsterdam & Philadelphia: John Benjamins
- Bauer, Laurie. 1997. Evaluative morphology: In search of universals. *Studies in Language* 21(3). 533–575.
- Booij, Geert. 2009. Compounding and Construction Morphology. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*, 201–216. Oxford: Oxford University Press.
- Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.
- Cappelle, Bert, Pascal Denis & Manuela Keller. 2018. Facing the facts of fake: a distributional semantics and corpus annotation approach. *Yearbook of the German Cognitive Linguistics Association* 6(1). 9–42.

- Dressler, Wolfgang Ulrich & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics. Diminutives and intensifiers in Italian, German, and other languages*. Berlin & New York: Mouton de Gruyter.
- Dressler, Wolfgang Ulrich & Annemarie Karpf. 1995. The theoretical relevance of Pre- and Protomorphology in language acquisition. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology*, 99–122. Dordrecht: Kluwer.
- Eitelmann, Matthias & Dagmar Haumann. 2019. Extravagant morphology. Workshop organized at the 52nd Annual Meeting of the Societas Linguistica Europaea (SLE 2019), Leipzig University, 21–24 August 2019.
- Firth, John Rupert. 1957. A synopsis of linguistic theory 1930–1955. In John Rupert Firth (ed.), *Studies in linguistic analysis*, 1–32. Oxford: Blackwell.
- Geurts, Bart. 2000a. Explaining grammaticalization (the standard way). *Linguistics* 38(4). 781–788.
- Geurts, Bart. 2000b. Function or fashion? Reply to Martin Haspelmath. *Linguistics* 38(6). 1175–1180.
- Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goublomme, Florette. 2019. *Het gebruik van het adjectief “fake” in het Nederlands. Corpusstudie en enquête bij moedertaalsprekers Nederlands over het gebruik van “fake” in geschreven en gesproken taal* [The use of the adjective “fake” in Dutch. Corpus study and survey of native Dutch speakers on the use of “fake” in written and spoken language]. Bachelorpaper supervised by K. Van Goethem, Université catholique de Louvain. 20 p.
- Grandi, Nicola & Livia Körtvélyessy (eds). 2015. *Edinburgh Handbook of Evaluative Morphology*. Edinburgh: Edinburgh University Press.

- Gyselinck, Emmeline & Timothy Coleman. 2016. *Je dood vervelen of je te pletter amuseren? Het intensiverende gebruik van de pseudoreflexieve resultatiefconstructie in hedendaags Belgisch en Nederlands Nederlands* [To be bored to death or to enjoy oneself tremendously? The intensifying use of the pseudo-reflexive resultative construction in contemporary Belgian and Dutch Dutch]. *Handelingen: Koninklijke Zuid-Nederlandse Maatschappij voor Taal- en letterkunde en geschiedenis* 69. 103–136.
- Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37(6). 1043–1068.
- Hoffmann, Thomas & Graeme Trousdale (eds.). 2013. *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.
- Keller, Rudi. 1994. *Language Change: The Invisible Hand in Language*. London: Routledge.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: ten years on. *Lexicography* 1. 7–36.
- Körtvélyessy, Lívia. 2014. Evaluative morphology. In Rochelle Lieber and Pavol Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 296–316. Oxford: Oxford University Press.
- Levshina, Natalia. 2015. *How to do Linguistics with R. Data exploration and statistical analysis*. Amsterdam & Philadelphia: John Benjamins.
- Levshina, Natalia & Kris Heylen. 2014. A radically data-driven Construction Grammar: Experiments with Dutch causative constructions. In Ronny Boogaart, Timothy Coleman & Gijsbert Rutten (eds.), *Extending the scope of Construction Grammar* (Cognitive Linguistics Research 54), 19–46. Berlin & New York: Mouton de Gruyter.
- Mattiello, Elisa. 2013. *Extra-grammatical Morphology in English: Abbreviations, Blends, Reduplicatives, and Related Phenomena*. Berlin & Boston: Mouton de Gruyter.

- Maechler, Martin, Rousseeuw, Peter, Struyf, Anja, Hubert, Mia & Kurt Hornik. 2019. cluster: Cluster Analysis Basics and Extensions. R package version 2.0.9.
- Norde, Muriel. 2009. *Degrammaticalization*. Oxford: Oxford University Press.
- Norde, Muriel & Kristel Van Goethem. 2014. Bleaching, productivity and debonding of prefixoids. A corpus-based analysis of “giant” in German and Swedish. *Lingvisticae Investigationes* 37(2). 256–274.
- Norde, Muriel & Kristel Van Goethem. 2015. *Emancipatie van affixen en affixoiden: Degrammaticalisatie of lexicalisatie?* [Emancipation of affixes and affixoids: Degrammaticalization or lexicalization?]. *Nederlandse Taalkunde* 20(1). 109–148.
- Norde, Muriel & Kristel Van Goethem. 2018. Debonding and clipping of prefixoids in Germanic: Constructionalization or Constructional Change? In Geert Booij (ed.), *The Construction of Words* (Studies in Morphology 4), 475–518. Cham etc.: Springer.
- R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. URL <https://www.R-project.org/>
- Schäfer, Roland. 2015. Processing and querying large web corpora with the COW 14 architecture. In Piotr Bański, Hanno Biber, Evelyn Breiteneder, Marc Kupietz, Harald Lungen, Andreas Witt (eds.), *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora (CMLC-3)*, 28–34. Mannheim: Institut für Deutsche Sprache.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.
- Szymanek, Bogdan. 1988. *Categories and categorization in morphology*. Lublin: Redakcja Wydawnictw Katolickiego Uniwersytetu Lubelskiego.

- Van Goethem, Kristel & Hendrik De Smet. 2014. How nouns turn into adjectives. The emergence of new adjectives in French, English and Dutch through debonding processes. *Languages in Contrast* 14(2). 251–277.
- Van Goethem, Kristel & Philippe Hiligsmann. 2014. When two paths converge: debonding and clipping of Dutch *reuze* ‘lit. giant; great’. *Journal of Germanic Linguistics* 26(1). 31–64.
- Van Goethem, Kristel & Matthias Hüning. 2015. From noun to evaluative adjective: conversion or debonding? Dutch *top* and its equivalents in German. *Journal of Germanic linguistics* 27(4). 366–409.
- von Mengden, Ferdinand & Anneliese Kuhle. forthcoming. Recontextualization and language change. To appear in *Folia Linguistica Historica* 41.
- Wouden, Ton van der. 2019. nep-. *Taalportaal*. Retrieved from <https://taalportaal.org/taalportaal/topic/pid/topic-13998813294731184>. (accessed 05 December 2019).
- Zwicky, Arnold M. & Geoffrey K. Pullum. 1987. Plain morphology and expressive morphology. General session and parasession on grammar and cognition. In Jon Aske, Natasha Beery, Laura Michaelis & Hana Filip (eds.), *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 330–340. Berkeley: Berkeley Linguistics Society.

Online corpora and dictionaries

- Sketch Engine: <http://www.sketchengine.eu>
- Etymology online: <http://www.etymonline.com>
- Online Woordenboek der Nederlandse Taal: <http://gtb.inl.nl>
- Etymologiebank Nederlands: <http://www.etymologiebank.nl/>
- Van Dale woordenboek Nederlands: <https://www.vandale.nl/>

¹ The etymological information for the Dutch items is based on <http://www.etymologiebank.nl/>.

² Dutch has borrowed English *fake* ‘counterfeit; imitation’ in the second half of the 20th century (1965, <http://www.etymologiebank.nl/trefwoord/fake>). The origin of English *fake* is however unknown. It is attested in London criminal slang as adjective (1775, ‘counterfeit’), verb (1812, ‘to rob’), and noun (1851, ‘a swindle of persons’, 1888, ‘a swindler’), but is probably older (<https://www.etymonline.com/search?q=fake>). In present-day Dutch, *fake* is also attested in adjectival contexts, possibly under the influence of adjectival usage in English.

³ These schemas represent formal (left) and semantic (right) properties of constructions and the symbolic link between them (indicated by a double arrow). Angled brackets represent constructions (including constructions within constructions). The uppercase subscripts denote word classes and phrase types, and the lowercase subscripts are lexical indices. Uppercase italics are unspecified phonological strings.

⁴ Source: <https://vrttaal.net/taaladvies-spelling/koppelteken-aaneenschrijven-trema>

⁵ In (13a), *100%* does not take scope over the NP as a whole, but only relates to *namaak* (it is a 100% fake, not a 100% fake elephant), hence *namaak* can only be an attributive adjective.

⁶ The corpus is available, after registration, at <https://www.webcorpora.org/>

⁷ The (synthetic) comparative form *faker*, however, is less likely to be of direct English origin. In *enTenTen15* (Kilgariff et al. 2014), *faker* (tagged as adjective) occurs 87 times, whereas the analytic comparative *more fake* (tagged as adjective) has a frequency of 366.

⁸ See <https://www.sketchengine.eu/>

⁹ See <https://www.sketchengine.eu/nltenten-dutch-corpus/>

¹⁰ For *fake*, 22 out of 56 distinctive collexemes are English loans (39%); for *nep*, the proportion is 4 English loans out of 46 distinctive collexemes (only 9%).

¹¹ The compound *fopspeen* is highly lexicalized in Modern Dutch, but etymologically it refers to a “losse speen, waarop men zuigelingen laat zuigen om ze zoet te houden” (‘loose teat, on which infants are sucked to keep them sweet’).

(<http://gtb.inl.nl/iWDB/search?actie=article&wdb=WNT&id=M016832.re.7&lemmodern=fopspeen&domein=0&conc=true>)

¹² One anonymous reviewer suggests that compounds such as *fakeblond* are a later development in Dutch, which indeed is perfectly possible.

¹³ In a context like (31), the interpretation of *fake* may vary: *fake* may receive a privative (‘fictive friends’) or non-privative (‘not-authentic, hypocritical friends’) meaning. Whether this semantic opposition is correlated to the presence of inflection remains to be investigated in a follow-up study.

¹⁴ Potential productivity is computed by dividing the number of hapax legomena in the sample by the number of tokens in the sample (cf. Baayen 2009).