**The semantic structuring of minimizing constructions in present-day Netherlandic Dutch: a distribution-based cluster analysis**

**Margot Van den Heede & Peter Lauwers**

**Abstract**:

This paper examines the semantic structuring of a paradigm of 89 minimizers, i.e., nouns that reinforce sentential negation in present-day Netherlandic Dutch, such as *meter* 'meter' in *voor geen meter vertrouwen* 'not to trust for a meter'. Cosine distances are computed on the basis of the predicates the minimizers combine with in a sample of 100 tokens downloaded from the Dutch Web corpus 2014 (nlTenTen14) and clustered according to the Partitioning Around Medoids (PAM) algorithm into nine semantic clusters. The clusters largely correspond to semantic categories such as taboo terms or units of money. This suggests that, in general, minimizers belonging to the same semantic domain are combined with a similar (core) set of predicates. Based on the shared predicates per cluster, we detect signs of analogical attraction between minimizers or, conversely, competition. Crucially, low silhouette widths enable us to identify outliers in their respective clusters, for instance, minimizing nouns that exhibit signs of context expansion, as shown by their combination with semantically non-harmonious verbs. As such, this paper provides a synchronic snapshot of the semantic processes involved in (incipient) grammaticalization of minimizing nouns and, more in general, it illustrates how distributional semantics offers a heuristic to analyze the structure of a network of comparable micro-constructions.

**Keywords**: minimizing constructions, Netherlandic Dutch, cluster analysis

**1. Introduction**

The present article focuses on the semantic structuring of so-called minimizing nouns or minimizersin present-day Netherlandic Dutch (Hoeksema 2002), such as *stap* 'step' in *geen stap vooruitkomen* 'lit. not to move one step forward'. We may consider them, from a constructionist perspective (e.g., Goldberg 1995), (lexical) minimizing *constructions*, as they depend in this particular use on the presence of another fixed element of negation, viz. *geen,* and as they are to be combined with a predicate to express a reinforcing – often non-compositional – meaning. Three additional slots (with optional elements) may be added to this construction: a preposition before [*geen* + N] and a post-determiner and/or a modifier between *geen* and the minimizer, as illustrated by the examples below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| [preposition] | [*geen*] | [post-determiner] | [modifier] | [minimizer] | [predicate] |
| / | *geen* 'no' | *ene* 'one' | / | *bal* 'ball' | *interesseren* 'to be interested' |
| *voor* 'for' | *geen* 'no' | / | / | *meter* 'meter' | *vertrouwen*  'to trust' |
| / | *geen* 'no' | / | *droog* 'dry' | *brood* 'bread' | *verdienen*  'to earn' |

Table 1: The different slots of the minimizing construction

Semantically, minimizing nouns refer to small entities, values, and quantities. By negating this minimum, the sentential negation is being reinforced: if one understands 'not even a word' of what someone else is saying, then one does not understand them 'at all'*.* The minimizing nouns can be assigned to different semantic groups, based on their original referential meaning. In Van den Heede & Lauwers (2021), we discussed the implicature originally involved in this pattern, and we roughly divided the set of minimizers – our dataset contains 244 different types – into three major categories on the basis of their meaning: (a) nouns such as *snars* (etym. 'draught') and *zier* (etym. 'maggot') that only function as minimizers and that no longer have a referential meaning (e.g., *geen zier geven om* 'not to give a ZIER about'), (b) nouns that also occur outside the construction but that no longer have a referential meaning in the construction, for instance, taboo words, (e.g., *geen fuck doen* 'not to do a fuck'), and (c) nouns that seem to have kept their original referential meaning in the minimizing construction, as they refer to minimal values or quantities on a time, distance, or weight scale (e.g., *geen gram aankomen* 'not to gain a gram').

The first goal of this paper is a merely descriptive one: to provide a classification of each of these minimizers based upon their distributional profile, i.e., the range of predicates they combine with. This will allow us to check if the actual use of the set of minimizers mirrors the intuitive semantic classifications as proposed by Van den Heede & Lauwers (2021). The second goal is to investigate the links within and between the clusters of minimizing constructions, with a particular focus on their peripheral members and instances of analogical levelling or, inversely, competition/specialization among minimizers, showing the dynamics with the paradigm of constructions. Finally, and most importantly, the close examination of outliers and unexpected cluster assignments will lead to the identification of various semantic processes of (ongoing) extension beyond the initial lexical-semantic niche of the minimizers (e.g., units of distance applied to cognition predicates). Such host-class expansion is expected in the case of incipient grammaticalization.

We will tackle these research questions with a methodology based on the general principles of distributional semantics. The results are further analyzed with the Partitioning Around Medoids (PAM) clustering algorithm to identify distributional classes. As such, the predicates the minimizers combine with act as the key element of the distributional contexts in which they occur and allow us to classify the minimizers on a data-driven, distributional basis.

This paper is organized as follows. In Section 2, we discuss the data selection and we provide a brief overview of our dataset. Section 3 is concerned with the results of the cluster analysis. First, we present a summary of the results, before moving on to a detailed analysis of the different clusters. This analysis zooms in on the meaning of the minimizers, their token and type frequencies, as well as the predicates they are most frequently combined with. The main findings of the study are summarized in Section 4.

**2. Data selection and general overview of the construction**

The first step in our study was to create a 'master' list of nouns that function as minimizers in present-day Netherlandic Dutch. We combined different strategies to achieve this goal, such as a literature review and several exploratory corpus studies (cf. Van den Heede & Lauwers 2021, for a more detailed description). Next, we retrieved all the instances of each noun on our list in the Netherlandic part of the Dutch Web corpus 2014 (nlTenTen14), a web corpus available on Sketch Engine (cf. Jakubíček *et al*. 2013). This Netherlandic Dutch subcorpus consists of 1,955,672,191 tokens. The following search syntax was used: [lemma="geen|gene|genen|geene|geenen"] [] {0,2} [word="MINIMIZER"] within<s/>, taking into account the possible variation of the minimizing noun, such as diminutives and spelling variants. We downloaded the results and manually filtered and annotated up to 100 relevant tokens (for the most frequent minimizers). As the search syntax illustrates, we did not take into account any other negative polarity contexts in which minimizers might appear, such as the preposition *zonder* (*zonder ook maar een seconde te aarzelen* 'without hesitating for even a second', nlTenTen14 corpus). We also excluded examples in which the scope of the minimizer is not the predicate of the sentence but another element such as a noun (*geen druppel regen* 'not a drop (of) rain') or an adverb (*geen cent extra* 'not a cent extra').

The result is a dataset of 244 minimizers, which are combined with 641 different predicates, for a total number of 6224 tokens[[1]](#footnote-1). The bar chart in Figure 1 shows the token frequency distribution of the different minimizers. The figure actually shows two types of token frequency: (a) the actual token frequency for the minimizers with less than 100 tokens at the left of the dotted line and (b) the extrapolated token frequency[[2]](#footnote-2) for the minimizers with more than 100 tokens at the right of the dotted line (for which we only cleaned and annotated the first 100 relevant tokens).

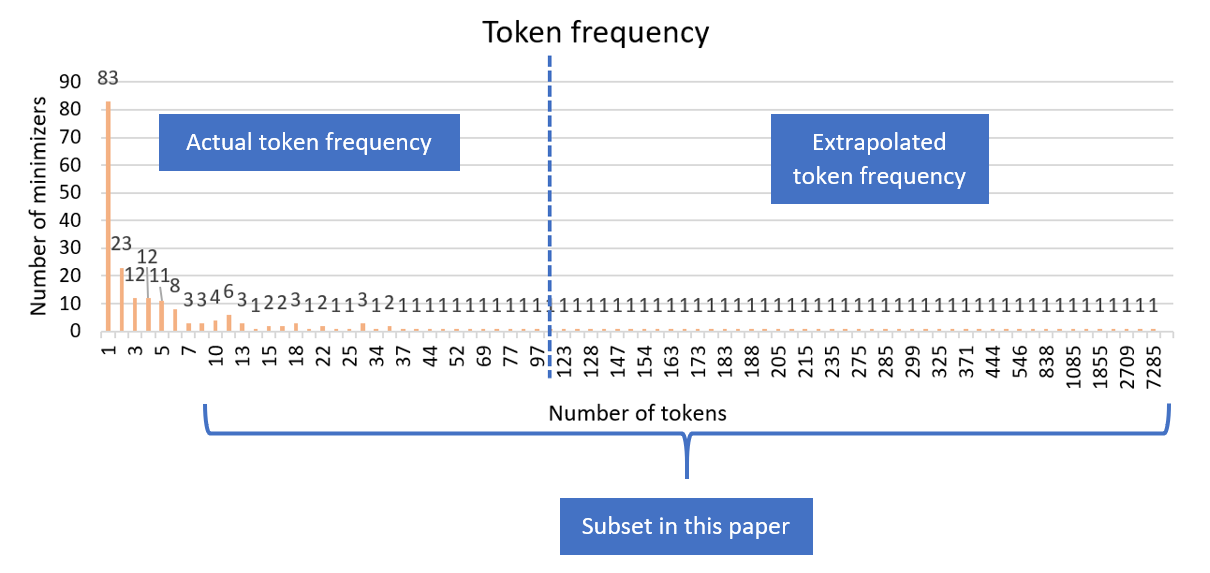


Figure 1: Token frequency distribution of the complete dataset (244 minimizers)

It is clear from Figure 1 that a lot of minimizers are very infrequent, with no less than 83 hapaxes, 23 dis legomena, and 12 tris legomena. In Van den Heede & Lauwers (2021) we already provided a qualitative analysis of these infrequent minimizers and argued that the minimizing construction has a wide lexical scope and a huge potential for growing, which is definitely a sign of the productivity (c.q. extensibility, Barðdal 2008) of the construction. In this article, we focus on a subset of the minimizers, namely the 89 minimizers with a token frequency of at least 10. The other minimizers are simply too infrequent to run a cluster analysis based on the predicates they combine with.

**3. Semantic structuring of the minimizing nouns**

In order to investigate the semantic structuring of the minimizers, we used cluster analysis. This bottom-up exploratory method clusters the minimizing nouns on the basis of the distribution of the predicates they occur with in the minimizing construction. Before analyzing the properties of the different clusters, we zoom in on the method that has been used.

*3.1 Method and terminology*

Our cluster analysis is based on a co-occurrence matrix with the minimizing nouns as rows and the predicates as columns. Each cell contains the co-occurrence frequency[[3]](#footnote-3). The idea behind this method is the widely accepted assumption that the meaning of a linguistic element can be inferred from the 'company it keeps' (Firth 1957: 11), that is, from its context words. Thus, we apply the general principle behind distributional semantics (akin to word embeddings; see i.a.Turney & Pantel 2010) to compute the semantic (dis)similarity between minimizers, yet on the sole basis of a context restricted to one particular slot (the predicate), very much like in the case of collostructional analysis (Stefanowitsch & Gries 2003).

On the basis of the table with the co-occurrence frequencies, we created a dissimilarity matrix in R (R version 4.3.1, using RStudio) with the dist\_cosine function of the svs package (Plevoets 2015). Cosine distance is generally used in distributional semantics because of its capacity to estimate the semantic distance between word vectors independent of their length, i.e., their total token frequency (cf. Heylen *et al*. 2015; Perek & Hilpert 2017). The values in the dissimilarity matrix range from 0 (closely related nouns) to 1 (very dissimilar nouns). This relatedness, which is much broader than mere synonymy, is thus based on the similarity of the distribution of each minimizing noun with its set of predicates. In the example below, we see, for instance, that *bal* and *biet* (as they are used in the minimizing construction) are more similar than *bal* and *barst*, because the cosine distance for the latter is higher.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *bal* 'ball' | *barst* 'crack' | *biet* 'beet' |
| *bal* 'ball' | 0 | 0.447 | 0.238 |
| *barst* 'crack' | 0.447 | 0 | 0.322 |
| *biet* 'beet' | 0.238 | 0.322 | 0 |

Table 2: Snapshot of the dissimilarity matrix

This dissimilarity matrix serves as the input for cluster analysis. In the linguistic literature, cluster analysis is a commonly used exploratory technique: it has been performed on the basis of cosine similarity resulting from co-occurrence matrices (Perek 2016; 2018), frequent collexemes (Gries & Stefanowitsch 2010), or behavioral profiles (Divjak & Gries 2006). Different methods exist to cluster data (see Divjak & Fieller 2014 for an overview of how the different methods of cluster analysis can be used in linguistic research). We used PAM, a type of non-hierarchical cluster analysis, which fits our needs. As a matter of fact, we are looking for maximally discrete semantic classes, without caring too much about the (chain-like) relations between these classes, which may cluster together in even bigger clusters. An additional benefit of this approach is that the output allows us to detect outliers within each cluster (with low silhouette values), indicating by the same token an alternative cluster assignment, since the nearest cluster is also identified. The PAM algorithm is a variant of k-means clustering (built around an existing medoid rather than a theoretical centroid, cf. Kaufman & Rousseeuw 2005: chapter 2). Since we did not have a particular hypothesis concerning the expected number of distributional clusters, we simply took the best solution (cf. below). Note that we used PAM rather than k-means because k-means is often not recommended if the distance measure is cosine[[4]](#footnote-4).

We performed the analysis in R by means of different packages: factoextra (Kassambara & Mundt 2020), fpc (Hennig 2020), and cluster (Maechler *et al*. 2019). The first step is to specify *k*, the desired number of clusters. To determine the ideal number of clusters, we used average silhouette[[5]](#footnote-5) as a diagnostic to assess the cluster quality of a given cluster solution: 'the silhouette value measures the degree of confidence in the clustering assignment of a particular observation', with values between 1 (well-clustered observations) and –1 (poorly clustered observations) (Divjak & Fieller 2014: 432). To calculate the silhouette width, the average intracluster distance is compared with the average intercluster distance to the nearest cluster (Rousseeuw 1987: 55–56).

Based on a comparison of the average silhouette width (ASW) values for *k* ranging from 2 to 10, the best option for our dataset appeared to be nine clusters, with an average silhouette width of 0.29[[6]](#footnote-6). For each of these (semantic) clusters, we determined also the shared predicates to illustrate their robustness. Crucially, cluster analysis does not only allow to cluster objects based on their similarity but also offers heuristics to detect outliers that have drifted away from (the core meaning of) a given semantic cluster of minimizers. More specifically, we will zoom in on those minimizers exhibiting unexpected class membership and cluster assignments with low silhouette values. The latter suggest that they are not really at home anymore in their semantic class. The outliers detected by the quantitative analysis will be subjected to qualitative scrutiny to identify the cases of semantic specialization or grammaticalization.

Recall that from a diachronic perspective, the grammaticalization process of minimizing nouns is a process including three stages which may be exemplified as follows (Breitbarth *et al*. 2013: 142–144):

(1) *I didn’t drink a drop*.

(2) *I couldn’t give a sausage*.

(3) *The problem didn’t detain us a jot*.

(examples from Breitbarth *et al*. 2013)

In the first stage, the minimizer is quite restricted and is often combined with 'verbs connected to its original meaning' (Breitbarth *et al*. 2013: 142). This stage is also mentioned by Hansen (2013: 56) with regard to the evolution of the French minimizers *pas* 'step' and *mie* 'crumb': ' […] prior to the situation observed in Old French, the reinforcing elements must have been confined to contexts equivalent to 'I don’t walk a step', 'I don’t eat a crumb', etc.’, in other words, to semantically compatible verbs. However, this stage could not be attested in Old French (Hansen & Visconti 2009). In the second stage, the scope of the minimizing noun is expanding, though the verbs are still transitive, as shown in (2). In the final stage, this 'transitivity restriction' is no longer present. The examples show that this grammaticalization process involves both semantic and syntactic aspects. Moreover, as examined by, among others, Hansen and Visconti (2009), certain discourse constraints may also apply. In this paper, we will provide a synchronic snapshot of the incipient grammaticalization of the Dutch minimizing nouns in stage 2. As such, following Himmelmann (2004: 32), we see grammaticalization, c.q. the emergence of a construction endowed with a grammatical function as 'a process of context expansion'. Himmelmann distinguishes between (a) host-class expansion, which refers to the expansion of the 'class of elements the gram is in construction with' (2004: 32) or 'an increase in construction type frequency' (Traugott & Trousdale 2013: 18), thus its productivity; (b) syntactic context expansion, and (c) semantic-pragmatic context expansion (for examples, see Himmelmann 2004: 32-33; Traugott & Trousdale 2013: 107). In this paper, only (a) and (c) will be relevant.

*3.2. Global overview of the semantic clusters*

The table below shows an overview of the different clusters, ranged from high to low average silhouette width (= ASW). We also provide silhouette widths for the minimizers within each cluster. In addition, we also indicate

- in bold: the medoids (= the item in the middle of the cluster)

- in italics: the minimizers with a silhouette value around 0

|  |  |  |
| --- | --- | --- |
|  | **Minimizers ≥ 10 tokens** | **ASW** |
| 1 | **seconde 'second'**(0.70), moment 'moment'(0.67), minuut 'minute'(0.64), dag 'day'(0.56), ogenblik 'instant'(0.50), tel 'moment; tick'(0.42) | 0.58 |
| 2 | **stuiver 'stuiver; five-cent piece'**(0.54), euro 'euro'(0.48), gulden 'guilder'(0.47), eurocent 'eurocent'(0.44), dubbeltje 'lit. double (two stuivers); ten-cent piece'(0.44), boterham 'sandwich'(0.42), brood 'bread'(0.41), duit 'duit'[[7]](#footnote-7)(0.40), cent 'cent'(0.38), kwartje 'little quarter; twenty-five cent piece'(0.37), dollar 'dollar'(0.33), *pepernoot 'ginger nut'(0.06)* | 0.39 |
| 3 | **fluit 'flute'**(0.56), moer 'MOER' (origin unclear, lit. 'nut; mother')(0.56), fuck 'fuck'(0.53), zak 'sack; scrotum'(0.53), bal 'ball'(0.52), kut 'cunt'(0.52), hol 'hole; arse'(0.50), reet 'arse'(0.49), donder 'thunder'(0.47), biet 'beet'(0.47), ruk 'jerk'(0.46), klap 'slap'(0.42), flikker 'faggot'(0.41), barst 'crack'(0.39), zier etym. 'maggot'(0.36), lor lit. 'rag'(0.34), kont 'bottom'(0.33), sikkepit etym. 'goat droppings'(0.31), pest 'plague'(0.30), (sode)mieter etym. 'mite'(0.28), snars etym. 'draught'(0.27), kloot 'ball; testicle'(0.23), hout lit. 'wood'(0.17), sier lit. 'show'(0.16), drol 'turd'(0.13), *bliksem 'lightning'(0.06)* | 0.38 |
| 4 | **spat 'spatter'**(0.59), komma 'comma'(0.50), steek 'sting; stab'(0.50), greintje 'lit. a little grain; not a bit of'(0.34), haar 'hair'(0.26), *jota 'iota'(-0.03)* | 0.36 |
| 5 | **slok 'sip'**(0.52), druppel 'drop'(0.49), drup 'drip'(0.34), *hap 'bite'(0.03)* | 0.34 |
| 6 | **wezen 'being, creature'**(0.39), kip 'chicken'(0.32), ziel 'soul'(0.32), sterveling 'mortal'(0.31), mens 'human'(0.31), kat 'cat'(0.29), hond 'dog'(0.20), glimp 'glimpse'(0.14), *korrel* '*granule; grain*'*(0.05)*, *vlieg 'fly'(0.05)*, *vin 'fin'(-0.03)* | 0.21 |
| 7 | kilo 'kilo'(0.30), **gram 'gram'**(0.29), *vinger 'finger'(0.002)*, *kruimel 'crumb'(-0.04)* | 0.14 |
| 8 | **zin 'sentence'**(0.32), regel 'line'(0.27), letter 'letter'(0.11), *woord 'word'(0.07)*, *toon 'tone'(0.06)*, *noot 'note'(0.04)*, *spaan 'chip (of wood)'(0.01)* | 0.13 |
| 9 | haarbreed 'hairbreadth'(0.20), **centimeter 'centimeter'(0.18)**, duimbreed 'thumbsbreadth'(0.18), *strobreed 'strawbreadth'(0.04)*, *stap 'step'(0.03)*, *millimeter 'millimeter'(0.02)*, *kilometer 'kilometer'(0.0008)*, *ding 'thing'(0)*, *schub 'scale'(-0.004)*, *meter 'meter'(-0.02)*, *nacht 'night'(-0.02)*, *spier 'muscle'(-0.04)*, *muis 'mouse'(-0.06)* | 0.04 |

Table 3: Overview cluster analysis

In general, most of the clusters, even those with low ASWs, (largely) correspond to certain semantic domains: 1) time, 2) money, 3) taboo words, 5) eating and drinking, 6) people and animals, 7) weight, 8) language and music, and 9) distance. The fourth cluster consists of defectors from other categories and, interestingly, illustrates that some minimizing nouns may show a similar distributional profile without sharing a common semantic core.

The table below provides some general information for each cluster: the number of minimizers it contains, the total number of distinct predicate types that are combined with this set of minimizers, as well as their total number of tokens. In order to compare the degree of distributional overlap between the minimizers in each cluster, we also computed the ratios between the total number of predicates and the total number of minimizers in each cluster, both in terms of types and tokens. It is interesting to note that cluster 3 (taboo words), by far the biggest cluster as to the number of minimizers and, by consequence, also the number of tokens, has the lowest number of predicates per minimizer-type/-token[[8]](#footnote-8). This means that the minimizers of cluster 3 form a robust cluster with a solid lexical base made up of a rather limited set of predicates, in contrast with, for instance, the minimizers of clusters 1 and 6. This finding will be analyzed more closely in the next part of the paper.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Cluster** | **# minimizers**  **\_types** | **# predicates** | **# predicates / # minimizers\_types** | **# minimizers**  **\_tokens** | **# predicates / # minimizers\_tokens** |
| 1 | 6 | 126 | 21 | 526 | 0.24 |
| 2 | 12 | 119 | 9.9 | 601 | 0.20 |
| 3 | 26 | 113 | 4.3 | 2018 | 0.06 |
| 4 | 6 | 81 | 13.5 | 375 | 0.22 |
| 5 | 4 | 45 | 11.3 | 223 | 0.20 |
| 6 | 11 | 191 | 17.4 | 772 | 0.25 |
| 7 | 4 | 37 | 9.3 | 217 | 0.17 |
| 8 | 7 | 73 | 10.4 | 435 | 0.17 |
| 9 | 13 | 129 | 9.9 | 696 | 0.19 |

Table 4: Frequency measures for each cluster

*3.3 Comparison of the different clusters*

In what follows, we analyze the different clusters in more detail. We provide a figure of the five most frequent predicates per cluster[[9]](#footnote-9) to make explicit the distributional basis on which the cluster is built. We also indicate the three most frequent predicates per minimizer. More than three predicates are listed if several predicates have the same token frequency. Note that sometimes we counted two variants as one single type (indicated by the pipe symbol |), such as *waard zijn*|*blijken* 'to be|to turn out to be worth', the semantically dominant element of the predicate being identical. For each minimizer, we also mention silhouette width, token frequency, and predicate type frequency, i.e., the number of different predicates that are combined with this minimizer in the dataset. For the minimizers with low silhouette values, we also refer to the closest (and sometimes more correct[[10]](#footnote-10)) cluster provided in the PAM output.

**Cluster 1: time**

Figure 2: Top predicates\_cluster1 (which cover 45.4% of the 526 tokens)

|  |  |  |  |
| --- | --- | --- | --- |
| **Minimizer** | **Silh. width** | **Token; type**  **frequency** | **Frequency top 3 predicates** |
| *dag* 'day' | 0.56 | 100; 43 | **1**.*vervelen* 'to be bored'\_(21), **2**.*missen* 'to miss'(14), **3**.*uit zijn gedachten zijn* 'to be out of his thoughts'(7) |
| *minuut* 'minute' | 0.64 | 100; 32 | **1**.*vervelen* 'to be bored'(32), **2**.*missen* 'to miss'(15), **3**.*slapen* 'to sleep'(7) |
| *moment* 'moment' | 0.67 | 100; 30 | **1**.*vervelen* 'to be bored'(37), **2**.*twijfelen* 'to doubt'(15), **3**.*bedenken* 'to consider'(7) |
| *ogenblik* 'instant' | 0.50 | 100; 49 | **1**.*twijfelen* 'to doubt'(11), **2**.*aarzelen* 'to hesitate'(7), **3**.*bedenken* 'to consider', *denken* 'to think', *vervelen* 'to be bored'(6) |
| *seconde* 'second' | 0.70 | 100; 35 | **1**.*vervelen* 'to be bored'(20), **2**.*twijfelen* 'to doubt'(15), **3**.*nadenken* 'to think'(10) |
| *tel* 'moment; tick' | 0.42 | 26; 19 | **1**.*stilzitten* 'to sit still'(4), **2**.*alleen laten* 'to leave alone', *in de steek laten* 'to abandon; to let down', *twijfelen* 'to doubt', *vervelen* 'to be bored'(2) |

Table 5: Overview minimizers cluster 1: silhouette width, token frequency, type frequency, and top predicates

According to the PAM algorithm, the best cluster in our dataset is composed of time minimizers. It is a semantically homogeneous category composed of six items. As the figure and the table above show, time minimizers are often combined with a whole bunch of semantically compatible verbs, often (but not always) endowed with a negative connotation, such as cognition verbs, e.g., *vervelen* 'to be bored', *twijfelen* 'to doubt', and *denken*; *nadenken*; *bedenken* 'to think'. Many predicates also refer to situations or feelings related to absence or isolation that are generally perceived as negative: *in de steek laten* 'to abandon', *er alleen voor staan* 'to stand alone', *misselijk*; *ziek*; *depressief zijn* 'to be nauseous; sick; depressed', *zich bang voelen* 'to feel scared', *zenuwachtig maken* 'to make nervous', and *saai zijn* 'to be boring'.

**Cluster 2: money**

Figure 3: Top predicates\_cluster2 (which cover 50.4% of the 601 tokens)

|  |  |  |  |
| --- | --- | --- | --- |
| **Minimizer** | **Silh. width** | **Token; type frequency** | **Frequency top 3 predicates** |
| *boterham* 'sandwich' | 0.42 | 44; 4 | **1**.*verdienen* 'to earn'(41), **2**.*door de keel krijgen* 'to get down the throat', *opleveren* 'to yield', *overhouden* 'to have left'(1) |
| *brood* 'bread' | 0.41 | 100; 2 | **1**.*verdienen* 'to earn'(99), **2**.*overhouden* 'to have left'(1) |
| *cent* 'cent' | 0.38 | 100; 31 | **1.***te makken hebben* 'to have money to spend'(16), **2.***betalen* 'to pay'(10), **3.***kosten* 'to cost'(9) |
| *dollar* 'dollar' | 0.33 | 12; 11 | **1**.*uitgeven* 'to spend'(2), **2.***betalen* 'to pay', *bezuinigen* 'to economize', *geven* 'to give', *hebben* 'to have', *missen* 'to miss', *ontvangen* 'to receive', *overhebben* 'to have left', *toevoegen* 'to add', *uitkeren* 'to pay out', *verdienen* 'to earn'(1) |
| *dubbeltje* 'ten-cent piece' | 0.44 | 39; 28 | **1**.*geven* 'to give', *overhouden* 'to have left', *verdienen* 'to earn'(3) |
| *duit* 'duit' | 0.40 | 12; 8 | **1**.*hebben* 'to have', *kosten* 'to cost', *verdienen* 'to earn', *waard zijn* 'to be worth'(2) |
| *euro* 'euro' | 0.48 | 100; 57 | **1.***uitgeven* 'to spend'(9), **2.***verdienen* 'to earn'(7), **3.***betalen* 'to pay', *waard zijn* 'to be worth'(5) |
| *eurocent* 'eurocent' | 0.44 | 35; 22 | **1.***verdienen* 'to earn', *uitgeven* 'to spend'(4), **2.***krijgen* 'to get', *opleveren* 'to yield'(3) |
| *gulden* 'guilder' | 0.47 | 12; 8 | **1.***waard zijn* 'to be worth'(3), **2.***uitgeven* 'to spend', *verdienen* 'to earn'(2) |
| *kwartje* 'quarter' | 0.37 | 10; 7 | **1.***waard zijn* 'to be worth'(4), **2.***geven* 'to give', *opleveren* 'to yield', *overhebben* 'to have left', *uitgeven* 'to spend', *verdienen* 'to earn', *zeilen* 'to sail'(1) |
| *pepernoot* 'ginger nut' | 0.06 | 37; 17 | **1.***snappen* 'to understand', *van bakken* 'neg.: to make a mess of; to suck', *verdienen* 'to earn'(5) |
| *stuiver* 'stuiver' | 0.54 | 100; 31 | **1.***waard zijn*|*blijken* 'to be|to turn out to be worth'(17), **2.***verdienen* 'to earn'(14), **3.***hebben* 'to have', *krijgen* 'to get'(8) |

Table 6: Overview minimizers cluster 2: silhouette width, token frequency, type frequency, and top predicates

The second cluster consists of units of money, including coins that are no longer in use (such as *dubbeltje*, *duit*, *kwartje*,and *stuiver*), with *euro* as the most productive minimizer. Most predicates, such as 'to earn'and'to be worth', are semantically harmonious verbs[[11]](#footnote-11):

(4) *Misschien is het bedrijf als zodanig geen euro meer waard, maar de omliggende landerijen vertegenwoordigen miljoenenkapitalen*.

'The company itself may not be worth a euro, but the surrounding land is worth millions.'

However, these same minimizer-predicate combinations can also be used metaphorically. For instance, in (5), the same predicate *waard zijn* is used in its metaphorical sense of 'be capable of':

(5) *Echt vervelend als je zo slecht slaapt. De nachten duren heel lang en overdag ben je op het laatst ook geen cent meer waard.*

'It's really annoying when you sleep so badly. The nights are very long and during the day at the end you're not worth a cent anymore.'

The following examples illustrate the next step in the process of context expansion of the minimizer, namely the combination with non-harmonious verbs, i.e., verbs that a priori have nothing to do with money, such as the verbs of indifference *uitmaken* 'to matter' (example 6) and *geven om* 'to care about' (example 7). These verbs form part of the most prototypical verbs of the minimizing construction; they belong to the top 20 most frequent predicates in the whole dataset. They mean something like 'attach a certain worth to' and are still somehow related to the idea of (monetary) value, since the concept 'worth' may be seen as a semantic generalization and abstraction starting from monetary value (cf. also in expressions such as *iets op prijs stellen* 'to find something very important', lit. 'to attach a price to something'). Note that in example (6), the link to money is also still present because the sentence is about the capital of a company. The hedge *letterlijk* recalls the link with the original referential context of the minimizer[[12]](#footnote-12):

(6) *Voor het vermogen of het resultaat van Ahold maakte de consolidatie letterlijk geen euro uit*.

'To the capital or the result of Ahold, the consolidation literally didn’t matter a euro.'

(7) *Luther gaf geen stuiver om de "onfeilbare leer van de Kerk", want daar geloofde hij helemaal niet in* […]

'Luther didn’t give a stuiver about the “infallible teaching of the Church”, because he didn’t believe in it at all […]'

Other examples of context expansion are given below. The verb *vertrouwen voor* 'to trust for' (example 8)is quite frequently used in this cluster (3x *cent,* 2x *dubbeltje*, 2x *eurocent*,and 3x *stuiver*), which points toward another extension of the semantic scope of the minimizer based on a metonymic association with a metaphorical flavor: the minimal amount of trust is metaphorically expressed in terms of the limited monetary value involved in a (fictive) monetary transaction that such trust would justify (= contiguity). Example (9), by contrast, shows the expansion of one specific micro-construction, namely the unique combination of the minimizer *kwartje* with the completely unrelated verb 'to sail', which is a hapax predicate in the entire dataset.

(8) *Ik vertrouw dit hele gedoe voor geen dubbeltje*…

'I don't trust this whole thing for a dime…'

(9) *Het zeilt voor geen kwartje. Ligt vrij snel plat, reageert nauwelijks op het roer en afstellen is ook een drama.*

'It doesn’t sail for a quarter. It flattens quite quickly, hardly responds to the rudder and adjusting is also a drama.'

Besides units of money, the second cluster also contains three nouns that at prima facie seem to refer to quantities of food but that quite conventionally symbolize the notion of 'earnings' in a metaphorical way: *brood* 'bread', *boterham* 'sandwich', and *pepernoot* 'ginger nut'. As a matter of fact, *bread* has become emblematic of 'income' in the nearly fixed V-N combination *zijn brood verdienen*, which means 'to earn one’s living'; cf. also the saying *de een zijn dood is de ander zijn brood* 'one man’s death is another man’s bread'. As such, these metaphorical extensions of the minimizer prefer the predicate *verdienen*. That is why they are situated in this second cluster. The third minimizer referring to food is *pepernoot* 'ginger nut', comparable to the metaphorical use of English 'peanuts'. This is the minimizer with the lowest silhouette value (0.06), which means that the clustering assignment is less good. Different groups of predicates are associated with *pepernoot*. First of all, three verbs are related to money, which explains why the minimizer is assigned to cluster 2: *verdienen* 'to earn', *opleveren* 'to yield', and *kosten* 'to cost'. In combination with these predicates, *pepernoot* is used symbolically for (absence of) money or value, as was the case for *brood* and *boterham*. Its low silhouette value is due to the fact that *pepernoot* is also combined – by extension – with verbs that have something to do with (un)successful actions and the low (monetary) value they subsequently generate, often in the context of sports: *raak schieten* 'to hit the target', *van bakken* (used in a negative context: 'to suck'), and *scoren* 'to score'. As such, these verbs are linked to the central (metaphorical) sense '*pepernoot* value' through a metonymic association between the events they denote and an effect, viz. earning money or, more generally, creating value. Third, by analogy to cluster 3 (i.e., the taboo cluster, which is the nearest neighbor of *pepernoot* according to our cluster analysis), the minimizer also occurs with several cognitive predicates such as *snappen*; *begrijpen* 'to understand' and *voorstellen* 'to mean'. *Snappen* and *begrijpen* also form part of the 20 predicates most frequently used in the entire dataset. We should keep in mind, however, that this top 20 is dominated by the predicate types of cluster 3 because this cluster has the highest number of minimizers, which, moreover, are often very token-frequent (cf. table 4 above).

In summary, it is quite obvious why these minimizers have been clustered together in cluster 2. They all refer to minimal monetary values and are subject to semantic generalization/abstraction and can combine with predicates that are linked to the realm of creating value through metonymic association. The core predicate of this cluster is *verdienen* 'to earn', which, crucially, is also part of the nearly fixed expressions with *brood* and *boterham*, which are conventional metaphors for 'earnings'.

**Cluster 3: taboo words**

Figure 4: Top predicates\_cluster3 (which cover 37.2% of the 2018 tokens)

|  |  |  |  |
| --- | --- | --- | --- |
| **Minimizer** | **Silh. width** | **Token; type frequency** | **Frequency top 3 predicates** |
| *bal* 'ball' | 0.52 | 100; 23 | **1.***snappen* 'to understand'(18), **2.***uitmaken* 'to matter'(13), **3.***interesseren* 'to be interested'(12) |
| *barst* 'crack' | 0.39 | 100; 23 | **1.***geloven* 'to believe'(25), **2.***interesseren* 'to be interested', *kunnen schelen* 'to care'(10) |
| *biet* 'beet' | 0.47 | 100; 29 | **1.***interesseren* 'to be interested'(26), **2.***snappen* 'to understand'(17), **3.***kunnen schelen* 'to care'(12) |
| *bliksem* 'lightning' | 0.06 | 22; 12 | **1.***aangaan* ‘to concern', *te maken hebben met* 'to have to do with', *zien* 'to see'(3) |
| *donder* 'thunder' | 0.47 | 100; 27 | **1.***uitmaken* 'to matter'(23), **2.***kunnen schelen* 'to care'(10), **3.***te maken hebben met* 'to have to do with'(9) |
| *drol* 'turd' | 0.13 | 100; 30 | **1.***kosten* 'to cost'(34), **2.***uitmaken* 'to matter'(11), **3.***snappen* 'to understand'(6) |
| *flikker* 'faggot' | 0.41 | 100; 40 | **1.***doen* 'to do'(12), **2.***uitvoeren* 'to do'(8), **3.***boeien* 'to care', *uitmaken* 'to matter'(7) |
| *fluit* 'flute' | 0.56 | 100; 36 | **1.***uitmaken* 'to matter'(15), **2.***interesseren* 'to be interested'(12), **3.***snappen* 'to understand', *zien* 'to see'(6) |
| *fuck* 'fuck' | 0.53 | 100; 36 | **1.***uitmaken* 'to matter'(9), **2.***doen* 'to do'(9), **3**.*interesseren* 'to be interested'(8) |
| *hol* 'hole' | 0.50 | 100; 29 | **1.***snappen* 'to understand'(16), **2.***interesseren* 'to be interested'(14), **3.***begrijpen* 'to understand'(10) |
| *hout* 'wood' | 0.17 | 100; 14 | **1.***kloppen* 'to be right'(36), **2.***snappen* 'to understand'(24), **3.***begrijpen* 'to understand'(17) |
| *klap* 'slap' | 0.42 | 100; 29 | **1.***aan vinden* 'to find it'(11), **2.***uitmaken* 'to matter'(10), **3.***aan zijn* 'to be to it; to like'(8) |
| *kloot* 'testicle' | 0.23 | 24; 17 | **1.***snappen* 'to understand'(5), **2.***verdienen* 'to earn'(3), **3.***aan hebben* 'to be of use'(2) |
| *kont* 'bottom' | 0.33 | 17; 12 | **1.***snappen* 'to understand'(3), **2.***doen* 'to do', *kosten* 'to cost', *uitmaken* 'to matter'(2) |
| *kut* 'cunt' | 0.52 | 35; 20 | **1.***doen* 'to do', *snappen* 'to understand'(4), **3.***interesseren* 'to be interested', *uitmaken* 'to matter'(3) |
| *lor* 'rag' | 0.34 | 52; 11 | **1.***interesseren* 'to be interested'(26), **2.***kunnen schelen* 'to care'(10), **3.***zich aantrekken* 'to care about, to be concerned about'(5) |
| *mieter* 'mite' | 0.28 | 71; 30 | **1.***weten* 'to know'(12), **2.***doen* 'to do'(6), **3.***aan hebben* 'to be of use'(5) |
| *moer* 'MOER' | 0.56 | 100; 35 | **1.***interesseren* 'to be interested'(11), **2.***uitmaken* 'to matter'(9), **3.***kunnen schelen* 'to care'(8) |
| *pest* 'plague' | 0.30 | 18; 11 | **1.***aangaan* 'to concern', *aan vinden* 'to find it', *aan zijn* 'to be to it; to like', *doen* 'to do', *helpen* 'to help', *uitmaken* 'to matter', *weten* 'to know'(2) |
| *reet* 'arse' | 0.49 | 100; 35 | **1.***doen* 'to do'(17), **2.***uitmaken* 'to matter'(10), **3.***kunnen schelen* 'to care'(8) |
| *ruk* 'jerk' | 0.46 | 100; 34 | **1.***doen* 'to do'(10), **2.***kunnen schelen* 'to care'(9), **3.***uitmaken* 'to matter'(8) |
| *sier* 'show' | 0.16 | 10; 8 | **1.***geven om* 'to care about', *helpen* 'to help'(2), **3.***interesseren* 'to be interested', *opknappen* 'to recover from it', *te maken hebben met* 'to have to do with', *uithalen* 'to be of use', *uitmaken* 'to matter', *vervelen* 'to be bored'(1) |
| *sikkepit* 'goat droppings' | 0.31 | 69; 30 | **1.***geloven* 'to believe'(9), **2.***begrijpen*, *snappen* 'to understand'(7) |
| *snars* 'draught' | 0.27 | 100; 21 | **1.***begrijpen* 'to understand'(28), **2.***snappen* 'to understand'(20), **3.***geloven* 'to believe'(17) |
| *zak* 'sack' | 0.53 | 100; 29 | **1.***snappen* 'to understand', *uitmaken* 'to matter'(12), **3.***aan vinden* 'to find it'*, doen* 'to do'(11) |
| *zier* 'maggot' | 0.36 | 100; 23 | **1.***interesseren* 'to be interested'(24), **2.***helpen* 'to help'(22), **3.***geven om* 'to care about'(15) |

Table 7: Overview minimizers cluster 3: silhouette width, token frequency, type frequency, and top predicates

As we mentioned above (§ 3.2), cluster 3 contains the biggest number of minimizers, mostly taboo nouns. It is quite probable that analogical attraction exists between them, based on their common source semantics (sexual, scatological)[[13]](#footnote-13). In addition, a considerable number also share formal features, such as identical initial letters (for instance, *bal*, *barst* and *flikker*, *fluit*) or limited word length (such as *bal*, *hol*, *lor*, *ruk*, and *zak*), a usual property of words with strong expressive and emotive function, according to one of our reviewers. Besides this semantic and formal analogy, taboo minimizers are united by their high degree of expressiveness and minimal referential semantics, which responds to the speakers’ need for 'extravagance' (Haspelmath 1999: 1055; Keller 1994). Recall that taboo nouns are also used in other constructions such as 'emphatic wh-questions' (e.g., 'where the fuck are you') (Hoeksema 2019: 6), 'rude imperatives' (e.g., 'fuck off') (Hoeksema 2019: 5), and 'the intensifying fake reflexive resultative construction' (e.g., *schrok me de tieten van me lijf af* 'startled myself the tits off my body off') (Gyselinck & Colleman 2018: 4).

As shown in the table above, the top three predicates of these taboo minimizing nouns are very similar and overlap to a great extent. Quite surprisingly, despite their lack of referential meaning, and hence, an expected absence of semantic combinatory restrictions, they are often combined with the same types of predicates such as cognition verbs (*snappen, begrijpen*…) and verbs of indifference (*uitmaken, kunnen schelen, interesseren*…). Due to the high number of (token-frequent) taboo minimizers and their similar distributional profiles, the predicates these minimizer combine with are the most frequent and hence the most prototypical predicates of the minimizing construction considered at the macro level. Again, this massive similar behavior can be explained by the mechanism of analogy: taboo minimizers connected through analogy also recruit in the same small pool of predicates. In this sense, our data confirm the statement of Postma (2001: 291) that taboo minimizers are 'virtually interchangeable'.

Next to the typical taboo words, cluster 3 also contains several other minimizing nouns that have lost their referential meaning, in general (e.g., *snars, zier*) or only in the minimizing construction (e.g., *hout*, *klap*). These bleaching minimizers have been attracted to the prototypical cluster, to the extent that their distributional behavior is very much like taboo minimizers. Hoeksema (2001) analyzed the diachronic evolution of the distributional profile of *snars* and *zier*. He concludes that during the 20th century, these minimizers 'show increasing semantic specialization, hence narrowing of their distribution'[[14]](#footnote-14) (Hoeksema 2001: 18): *snars* develops a preference for verbs of cognition whereas *zier* favors verbs of indifference. This tendency is confirmed by our synchronic dataset. Table 7, featuring the most frequent predicates, shows that the top three predicates of *snars* are verbs of cognition and that this is also the case with, for instance, *sikkepit.* On the other hand, the top three predicates of *zier*,together with, for example, *donder* and *lor*, confirm their preference for verbs of indifference.

Now, let us consider one minimizer in particular that has a somewhat deviating distributional profile: *drol* 'turd; excrement'. As is shown in the frequency spectrum below, this minimizeroccurs with four of the top five predicates of cluster 3 (*interesseren, snappen, uitmaken*,and *begrijpen*). Yet the most frequent combination is *geen drol kosten* (lit. 'not to cost a turd'), *drol* having taken the more general sense of something worthless (an evolution comparable to *pepernoot* in cluster 2):

(10) *Alhoewel ik toch steeds meer afgunstige gezichten krijg als ik vertel dat er een eenpersoonskamer met badkamertje en keukentje bij inbegrepen is, én het recht op huurtoeslag waardoor ik dus voor geen drol straks drie jaar als een prinsesje (voor studentenmaatstaven) op de Koestraat woon!*

'Although I do get more and more jealous faces when I tell them that it includes a single room with bathroom and kitchenette, as well as the right to rent allowance, which means I will soon be living (lit.) for no turd like a princess (by student standards) on the Koestraat for three years!'

So, *drol* appears to combine with predicates typical for taboo minimizers and for nouns referring to units of money, including also *waard zijn* 'to be worth' and *krijgen* 'to get', predicates typical of its nearest cluster (2). This in-between status of *drol* is also reflected in its low silhouette value (0.13).

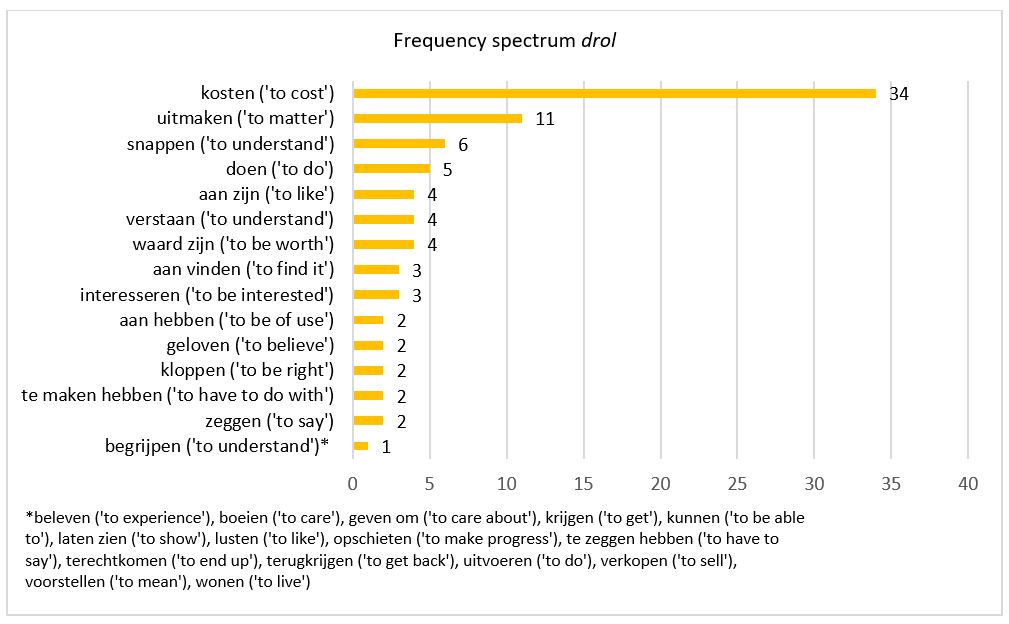


Figure 5: Frequency spectrum *drol*

As already suggested by *drol,* the taboo minimizers of cluster 3 are not completely unrelated to *kosten* 'to cost', which belongs to the top 20 most frequent predicates. Yet, the frequencies are much lower (1x *biet*, 1x *flikker*, 3x *fluit*, 2x *kont*, 2x *(sode)mieter*, 2x *moer*, 1x *reet*, and 5x *ruk*).

**Cluster 4: defectors from other categories**

Figure 6: Top predicates\_cluster4 (which cover 59.5% of the 375 tokens)

|  |  |  |  |
| --- | --- | --- | --- |
| **Minimizer** | **Silh. Width** | **Token; type frequency** | **Frequency top 3 predicates** |
| *greintje* 'a little grain' | 0.34 | 12; 10 | **1.+2.***ontgaan* 'to fail to notice', *veranderen* 'to change'(2), **3.***afdoen aan* 'to alter', *bang zijn* 'to be afraid', *bezitten* 'to possess', *nodig hebben* 'to need', *verhogen* 'to raise', *verschillen* 'to differ', *vinden* 'to find', *zakken* 'to drop'(1) |
| *haar* 'hair' | 0.26 | 51; 20 | **1.***schelen* 'to make a difference'(17), **2.***veranderen* 'to change'(8), **3.***denken* 'to think', *verroeren* 'to stir; to move', *vertrouwen* 'to trust'(3) |
| *jota* 'iota' | -0.03 | 100; 29 | **1.***snappen* 'to understand'(24), **2.***begrijpen* 'to understand', *veranderen* 'to change'(14) |
| *komma* 'comma' | 0.50 | 12; 6 | **1.***veranderen* 'to change'(5), **2.***liegen* 'to lie', *staan* 'to stay'(2) |
| *spat* 'spatter' | 0.59 | 100; 24 | **1.***veranderen* 'to change'(74), **2.***opschieten* 'to make progress', *terechtkomen* 'to end up', *uitmaken* 'to matter'(2) |
| *steek* 'sting; stab' | 0.50 | 100; 17 | **1.***veranderen* 'to change'(42), **2.***opschieten* 'to make progress'(17), **3.***zien* 'to see'(14) |

Table 8: Overview minimizers cluster 4: silhouette width, token frequency, type frequency, and top predicates

The fourth cluster unites minimizers that seem to originate from multiple semantic sources. Their main link is the frequent association with *veranderen* 'to change', which is used in 145 out of the 375 tokens of this cluster (= 39% of the instances). However, this predicate unites at least two different senses. In combination with *greintje, haar, spat*,and *steek*, the verb is intransitive and means 'to become different'. When *veranderen* is combined with *jota* and *komma*, by contrast, it is transitive and means 'to make different' or 'to change'. For instance, *geen jota veranderen* is often used in (archaic) contexts of religion and law: to change not a jot of the Bible/Thora/the law[[15]](#footnote-15). This polysemy of *veranderen* is not detected by our analysis, since we clustered the minimizers at type level and not at sense level.

(11) *Beste Joos, zo te zien ben je geen spat veranderd in de afgelopen zestig jaar*.

'Dear Joos, it seems you have not changed a spatter in the last sixty years.'

(12) *Met de nieuwe jeugdwet verandert er geen jota aan de onveilige situatie of omstandigheden voor de jeugd.*

'The new youth act will not change one iota to the unsafe situation or conditions for the youth.'

Note that the minimizers initially related to language, *komma* 'comma' and *jota* 'iota', are separated from the typical language minimizers (*letter* 'letter', *regel* 'line', *woord* 'word',and *zin* 'sentence'), which belong to cluster 8. *Komma*, however, has still cluster 8 as its nearest neighbor, while *jota* is more oriented toward cluster 4, the taboo minimizers, because it is combined with different types of predicates (similar to *pepernoot* [cluster 2] *and drol* [cluster 3]).

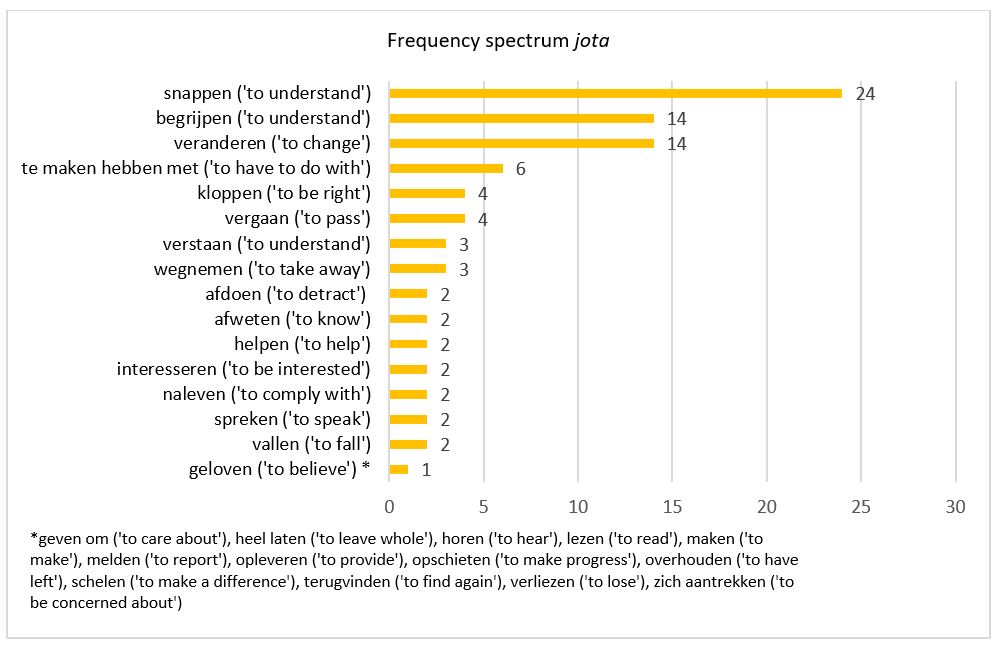


Figure 7: Frequency spectrum *jota*

According to its original referential meaning, *jota,* the smallest letter in the Greek alphabet, is related to language. It is therefore combined with verbs that are clearly related to oral and written communication (*spreken* 'to speak'*, horen* 'to hear'*,* and *lezen* 'to read') or, more generally, in contexts related to language, for instance, with cognition verbs:

(13) *Het begint namelijk zeer theoretisch met de uitleg van het Thaise alfabet en ik snap er geen jota van.*

'It starts off very theoretically with an explanation of the Thai alphabet and I don't understand an iota of it.'

However, a lot of speakers no longer have the etymological meaning of *jota* in mind, it easily expands to contexts beyond language. Again, the prototypical predicates of the minimizing construction (i.e., the predicates of cluster 3) pop up, such as *helpen* 'to help' (example 14), *zich aantrekken van* 'to be bothered' (example 15), and other verbs of indifference (*interesseren* 'to be interested', *geven om* 'to care about').

(14) *Het helpt geen iota.*

'It doesn’t help an iota.'

(15) *Kijk uit het raam en zie hoe de levende natuur zich geen jota aantrekt van het lawaai van de voorbijsnellende treinen.*

'Look out of the window and see how living nature is not bothered an iota by the noise of passing trains.'

In these examples, we clearly see that *jota* evolves toward a noun without any referential meaning, like the taboo minimizers of cluster 3. Although it is attracted to the prototypical instances of the construction, in contrast to *zier*, *hout*…, it does not yet form part of cluster 3 because the referential meaning is still present in several instances, for example, in the context of religion. In sum, the distributional profile of *jota* is a combination of verbs that are used in the context of religion/law (*veranderen* 'to change', *vergaan* 'to pass'), verbs related to language, and verbs that are typical of the taboo minimizers. This explains its negative silhouette width (–0.03).

**Cluster 5: eating and drinking**

Figure 8: Top predicates\_cluster5 (which cover 74.4% of the 223 tokens)

|  |  |  |  |
| --- | --- | --- | --- |
| **Minimizer** | **Silh. width** | **Token; type frequency** | **Frequency top 3 predicates** |
| *drup* 'drip' | 0.34 | 10; 9 | **1.***drinken* 'to drink'(2), **2.***komen in* 'to get in', *kunnen bij* 'to add', *lekken* 'to leak', *missen* 'to miss', *regenen* 'to rain', *verliezen* 'to lose', *zijn over* 'to be left', *zitten in* 'to be in it'(1) |
| *druppel* 'drop' | 0.49 | 100; 31 | **1.***drinken* 'to drink'(49), **2.***uit komen* 'to come out'(7), **3.***morsen* 'to spill'(6) |
| *hap* 'bite' | 0.03 | 100; 11 | **1.***door de keel krijgen*|*gaan* 'to get down the throat'(74), **2.***eten* 'to eat'(14), **3.***binnenkrijgen* 'to swallow'(3) |
| *slok* 'sip' | 0.52 | 13; 4 | **1.***drinken* 'to drink'(8), **2.***ophebben* 'to have had'(3), **3.***binnenhouden* 'to keep it in', *door de keel krijgen* 'to get down the throat'(1) |

Table 9: Overview minimizers cluster 5: silhouette width, token frequency, type frequency, and top predicates

Cluster 5 is composed of nouns that relate to the activity of eating and drinking. However, despite this common base, the number of shared predicates is not that high. This is partly due to the low token frequency of *drup* and *slok*, which is related to the fact that the number of predicates that belong to this semantic niche of eating and drinking is quite restricted. *Slok* 'sip' always refers to drinking, in contrast to *druppel* 'drop' and *drup* 'drip', which also function as minimal amounts of rain (16a), sweat (16b), and other liquids, which considerably extends the number of semantically compatible events (and hence predicates) of *druppel*.

(16) a) *Het had daar toen al drie jaar geen drup geregend.*

'It hadn't rained a drop there for three years.'

b) *Een boomlange man, in een zwart leren pak, die desondanks de hele dag nog geen druppel zweette.*

'A tall man, in a black leather suit, who nevertheless did not sweat a drop all day.'

*Hap* 'bite'has some ingestion predicates in common with *druppel* (*binnenkrijgen* 'to swallow', *proeven* 'to taste') and *slok* (*door de keel krijgen* 'to get down the throat'), but it has a low silhouette value (0.03) because it is the only minimizer that refers to ingested non-liquid food, which is reflected in the list of predicates. Moreover, *geen hap door de keel krijgen* is a very frequent combination, almost a fixed expression.

**Cluster 6: people and animals**

Figure 9: Top predicates\_cluster6 (which cover 53.5% of the 772 tokens)

|  |  |  |  |
| --- | --- | --- | --- |
| **Minimizer** | **Silh. width** | **Token; type frequency** | **Frequency top 3 predicates** |
| *glimp* 'glimpse' | 0.14 | 15; 5 | **1.***zien* 'to see'(6), **2.***missen* 'to miss'(5), **3.***laten zien* 'to show'(2) |
| *hond* 'dog' | 0.20 | 100; 50 | **1.***zijn* 'to be'(11), **2.***komen* 'to come'(9), **3.***lezen* 'to read'(8) |
| *kat* 'cat' | 0.29 | 34; 25 | **1.***zien* 'to see', *zijn* 'to be'(3), **3.***gebruiken* 'to use', *kijken* 'to look', *kwaad doen* 'to harm', *lezen* 'to read', *zich aantrekken* 'to care about; to be concerned about'(2) |
| *kip* 'chicken' | 0.32 | 100; 29 | **1.***zijn* 'to be'(28), **2.***bekennen* ‘to see; to detect', *zien* 'to see'(11) |
| *korrel* 'granule; grain' | 0.05 | 14; 14 | **1.***afbreken* 'to break off', *afnemen* 'to decrease', *bij komen* 'to add', *bij kunnen* 'can be added', *gunnen* 'to grant', *morsen* 'to spill', *ontvangen* 'to receive', *over zijn* 'to be left', *rapen* 'to pick up', *redden* 'to save', *stuiven* 'to drift', *verliezen* 'to lose', *vinden* 'to find', *zien* 'to see'(1) |
| *mens* 'human' | 0.31 | 100; 66 | **1.***zien* 'to see'(12), **2.***bekennen* 'to see; to detect'(5), **3.***in leven laten* 'to keep alive', *weten* 'to know'(4) |
| *sterveling* 'mortal' | 0.31 | 100; 64 | **1.***bekennen* 'to see; to detect'(13), **2.***zien* 'to see'(10), **3.***weten* 'to know'(5) |
| *vin* 'fin' | -0.03 | 84; 2 | **1.***verroeren* 'to stir'(83), **2.***bewegen* 'to move'(1) |
| *vlieg* 'fly' | 0.05 | 100; 2 | **1.***kwaad doen* 'to harm'(99), **2.***bekennen* 'to see; to detect'(1) |
| *wezen* 'being; creature' | 0.39 | 25; 8 | **1.***zien* 'to see'(10), **2.***bekennen* 'to see; to detect'(8), **3.***tegenkomen* 'to come across'(2) |
| *ziel* 'soul' | 0.32 | 100; 44 | **1.***bekennen* 'to see; to detect'(27), **2.***zien* 'to see'(9), **3.***weten* 'to know'(8) |

Table 10: Overview minimizers cluster 6: silhouette width, token frequency, type frequency, and top predicates

The sixth cluster is mainly composed of minimizers that are used to refer to an animate/human referent, instead of 'nobody'. Note that some of the minimizers, especially *mens* and *sterveling*, have a very high type frequency. The core predicates of this cluster are related to the visual domain, namely seeing someone or being present somewhere: *bekennen* 'to see; to detect', *zien* 'to see', *zijn* 'to be', *komen* 'to come', and the less frequent *tegenkomen* 'to meet; to come across', *vinden* 'to find', and *bespeuren* 'to be seen':

(17) *Ook wel fijn zo vroeg, heerlijk als de zon net opkomt en er geen kip te bekennen is op het water.*

'It is also nice this early, when the sun is just rising and there is not a chicken to be seen on the water.'

The minimizers with the lowest silhouette values are *vlieg* 'fly' (0.05), *vin* 'fin' (–0.03), and *korrel* 'granule; grain' (0.05). *Vlieg* and *vin* are part of (nearly) fixed expressions: *geen vlieg kwaad doen* 'not to harm; hurt a fly' and *geen vin verroeren* 'not to move a fin'. Logically, these token-frequent combinations make them highly idiosyncratic within the dataset, which strongly affects their clustering: even the most optimal cluster still exhibits only a very tiny link with the minimizer because, besides the idiosyncratic combination, there is not much room in the sample for predicates shared with other potential clusters. For instance, *vlieg* is only related to this cluster on the basis of a very small yet double link. First, because, by analogy, the same low-level image of 'harming a (banal) living creature' extends to *hond, kat, kip, wezen*,and *ziel,* whichare also combined with *kwaad doen* in the dataset, though only once or twice. The second link lies in the verb *bekennen* 'to see; to detect', frequently used with the other minimizers of this cluster, yet only once with *vlieg*. In the case of *vin*, the link with this cluster is even smaller, viz. the fact that its frequent predicate *verroeren* 'to stir; to move' appears to occur once with another minimizer, *wezen* 'being; creature'. In sum, although a priori *vlieg* and *vin* match with the other minimizers of this cluster (i.e., animal [parts]), they do not share the core predicates of the cluster. They only constitute peripheral extensions of the category, a phenomenon reminiscent of Lakoff’s chaining principle (1987). They do not share the prototypical basis of the category but are only related to it by means of small conceptual links with other individual members of the cluster.

The third peripheral member of this cluster, *korrel*,has a low silhouette value because of a completely different reason. As the table above reveals, none of its predicates occurs more than once. This is due to the fact that *korrel* can be used as a minimal amount of a lot of different things (in a similar vein as *druppel* above, cf. cluster 5): food (such as sugar, salt, grain…) (example 18a), sand (example 18b), or plastic (example 18c):

(18) a) *Toch is de gemiddelde zoutinname de afgelopen drie jaar met geen korrel afgenomen.*

'Yet average salt intake has not decreased by a grain over the past three years.'

b) *Een stuivend duin verplaatst alleen maar zand, er komt geen korrel bij.*

'A drifting dune only moves sand, not a grain is added.'

c) *Er gaat binnen Plasthill geen korrel verloren* […]

'No grain goes to waste at Plasthill […]'

Four of its predicates (*afbreken* 'to break off', *rapen* 'to pick up', *redden* 'to save', and *stuiven* 'to drift') are even hapaxes in the entire dataset. Because of its unique distributional profile, *korrel* is an example of a minimizer that is hard to cluster: it does not really belong to any cluster at all.

**Cluster 7: weight**

Figure 10: Top predicates\_cluster7

|  |  |  |  |
| --- | --- | --- | --- |
| **Minimizer** | **Silh. width** | **Token; type frequency** | **Frequency top 3 predicates** |
| *gram* 'gram' | 0.29 | 100; 21 | **1.***aankomen* 'to put on weight'(42), **2.***afvallen* 'to lose weight'(21), **3.***bijkomen* 'to gain weight'(8) |
| *kilo* 'kilo' | 0.30 | 25; 9 | **1.***aankomen* 'to put on weight'*, afvallen* 'to lose weight'(6), **3.***af zijn* 'to have lost', *bijkomen* 'to gain weight'(3) |
| *kruimel* 'crumb' | -0.04 | 26; 11 | **1.***overblijven* 'to be left; to remain'(10), **2.***over zijn* 'to be left'(4), **3.***achterblijven* 'to be left behind', *eten* 'to eat', *overlaten* 'to leave'(2) |
| *vinger* 'finger' | 0.002 | 65; 6 | **1.***aanraken* 'to touch'(57), **2.***verroeren* 'to stir; to move'(4), **3.***aankomen* 'to touch', *aan kunnen* 'to reach', *toegeven* 'to give in', *uitvoeren* 'to do'(1) |

Table 11: Overview minimizers cluster 7: silhouette width, token frequency, type frequency, and top predicates

The seventh cluster is a small cluster containing four minimizers. The two units of weight *kilo* and *gram* are, unsurprisingly, most often combined with gaining or losing weight (example 19a). Metaphorically, its scope also expands to abstract notions, both with verbs that are still linked to gaining/losing measurable quantities, but also way beyond, as shown by non-harmonious verbs such as *veranderen* 'to change' (example 19b) and *terecht zijn* 'to be justified' (example 19c). The examples illustrate that the minimizer evolves from indicating the lowest measurable quantity (in processes that involve measurable quantities) to the lowest degree of realization or truth of almost any kind of process.

(19) a) *Volgens mij ben ik door dit geneesmiddel enorm bijgekomen en ik kan geen enkele gram meer verliezen wat ik ook doe!*

'I think this medicine has made me gain a lot of weight and I cannot lose a single gram no matter what I do!'

b) *Maar het karakter van Den Dolder is geen grammetje veranderd.*

'But the character of Den Dolder has not changed a gram (diminutive).'

c) *Schuld die zelfs voor geen grammetje terecht is.*

'Guilt that is not even justified for a gram (diminutive).'

The other two minimizers in this cluster, *vinger* 'finger' and *kruimel* 'crumb', are not semantically related to *kilo* and *gram* and are peripheral members of the cluster with low silhouette values (0.002 and –0.04), which suggests bad or even wrong cluster assignment. As a matter of fact, *vinger* has a very idiosyncratic behavior and constitutes a highly conventionalized combination with *aanraken* ‘to touch’, a predicate that only occurs with *vinger* in the dataset. The only link with the weight minimizers is its hapax *aankomen* (a situation which reminds us of *geen vlieg bekennen*, cf. above). However, even this tiny link is actually spurious, since in combination with *vinger*, *aankomen*, it does not mean 'to gain weight' but 'to touch' (which is a synonymic expression of the frequent combination *met geen vinger aanraken* 'not to touch with a finger'). As already mentioned, polysemy (or homonymy) is not taken into account here. According to the algorithm, the next closest cluster of *vinger* is cluster 6, which results from the fact that *vinger* occurs with the verb *verroeren* 'to stir; to move', just like *vin*. This would have been a better cluster assignment for this highly conventionalized minimizer.

As to *kruimel*, this minimizer is used with predicates belonging to two semantic domains. On the one hand, it is often used to express that nothing, not even a crumb, is left over:

(20) *Ondertussen hebben Ellis en Pieter in het park voor een heerlijke Spaanse lunch gezorgd, er blijft geen kruimel over.*

'In the meantime, Ellis and Pieter have provided a delicious Spanish lunch in the park, not a crumb is left over.'

This idea can also be expressed by referring to a minimal unit of weight such as *gram,* which explains why *kruimel* is situated in this cluster. *Kruimel* and *gram* share two predicates: *achterblijven* 'to be left behind' and *verliezen* 'to lose', *achterblijven* being only combined with these two minimizers. On the other hand, *kruimel*, a minimal amount of food,is semantically related to the nouns of the fifth cluster (its nearest neighbor) that refer to eating and drinking activities, and they share several predicates, such as *eten* 'to eat', *verspillen* 'to waste', and *over zijn* 'to be left'. The two semantic domains to which *kruimel* is related attest to the process of semantic generalization undergone by this minimizer, from a minimal quantity of a particular kind of food (bread) to a minimal weight or even minimal quantity in general.

**Cluster 8: language and music**

Figure 11: Top predicates\_cluster8 (which cover 47.6% of the 435 tokens)

|  |  |  |  |
| --- | --- | --- | --- |
| **Minimizer** | **Silh. width** | **Token; type frequency** | **Frequency top 3 predicates** |
| *letter* 'letter' | 0.11 | 100; 28 | **1.***op papier hebben*|*komen*|*krijgen*|*staan*|*zetten* 'to have|come|get|be|put on paper(30), **2.***lezen* 'to read'(21), **3.***schrijven* 'to write'(10) |
| *noot* 'note' | 0.04 | 97; 24 | **1.***spelen* 'to play'(30), **2.***zingen* 'to sing'(19), **3.***horen* 'to hear'(14) |
| *regel* 'line' | 0.27 | 15; 7 | **1.***op papier krijgen*|*zetten* ' to get|put on paper'(5), **2.***schrijven* 'to write'(4), **3.***op het scherm komen* 'to appear on the screen'(2) |
| *spaan* 'chip (of wood)' | 0.01 | 100; 6 | **1.***heel blijven*|*houden*|*laten*|*zijn* 'to remain|keep|let|be whole'(87), **2.***overblijven* 'to remain', *overlaten* 'to leave'(5) |
| *toon* 'tone' | 0.06 | 10; 5 | **1.***uit* *komen* 'to come out'(4), **2.***uitkrijgen* 'to get out'(3), **3.***horen* 'to hear', *missen* 'to miss', *onderdoen* 'to be inferior to'(1) |
| *woord* 'word' | 0.07 | 100; 24 | **1.***reppen* *over* 'to talk about'(22), **2.***zeggen* 'to say'(16), **3.***liegen* 'to lie'(13) |
| *zin* 'sentence' | 0.32 | 13; 7 | **1.***schrijven* 'to write'(4), **2.***op* *papier* komen|*krijgen* 'to come|get on paper'(3), **3.***uit* *komen* 'to come out'(2) |

Table 12: Overview minimizers cluster 8: silhouette width, token frequency, type frequency, and top predicates

The predicates that are combined with the minimizers of cluster 8 are in line with the referential meaning of the minimizers for music (*toon* 'tone' and *noot* 'note') and language (*zin* 'sentence', *regel* 'line', *letter* 'letter', and *woord* 'word'). These two categories share indeed some verbs, as exemplified below:

(21) *Twintig jaar lang wist Stewart geen noot op papier te zetten na een confronterend gesprek met zijn toenmalige platenmaatschappij.*

'For twenty years, Stewart was unable to put a note to paper after a confrontational conversation with his then record label.'

(22) *Daarna peins ik verder over mijn nieuwe boek, maar er komt geen letter op papier.*

'Then I go on thinking about my new book, but not a letter gets put to paper.'

As to the language-related minimizers, we see some semantic specialization. A distinction can be drawn between *zin*, *regel*,and *letter* on the one hand, which are – unsurprisingly – more oriented toward written communication, and *woord* on the other hand, which is more into oral communication. Moreover, *regel* ('line, e.g., of code') is also combined with verbs related to programming language.

*Spaan* 'chip' constitutes the outlier of this category (sw : 0.01). Outside the construction, *spaan* refers to a small piece of wood, but this meaning is lost in the minimizing construction. The only link between *spaan* and its cluster is the fact that the medoid, *zin*, is also combined (but only once) with the verb *heel laten*, the most frequent predicate of *spaan*. Once again, the chain is extremely fragile.

(23) *Daar vertelde Bas van de Haterd over zijn eerdere boeken. Hoe de uitgever bijna geen zin heel had gelaten van zijn conceptversie.*

'There, Bas van de Haterd talked about his earlier books. How the publisher had left almost not a sentence intact from his draft version.'

*Spaan* is another example of a minimizer that is hard to cluster because it is very often combined with the same verb that is almost exclusively used with this minimizer. It has been attached to cluster 8, but cluster 7 would also have been an option, since *spaan* and *kruimel* share two predicates.

**Cluster 9: distance**

Figure 12: Top predicates\_cluster9

|  |  |  |  |
| --- | --- | --- | --- |
| **Minimizer** | **Silh. width** | **Token; type frequency** | **Frequency top 3 predicates** |
| *centimeter* 'centimeter' | 0.18 | 100; 48 | **1.***toegeven* 'to give in'(13), **2.***wijken* 'to give in; to give way'(12), **3.***vooruitkomen* 'to move forward'(9) |
| *ding* 'thing' | 0 | 12; 1 | **1.***bezorgd* *maken*|*zijn* 'to make, be worried'(12) |
| *duimbreed* 'thumbsbreadth' | 0.18 | 100; 7 | **1.***toegeven* 'to give'(76), **2.***wijken* 'to give in; to give way'(12), **3.***in* *de* *weg* *leggen*|*liggen* 'to get in the way'(6) |
| *haarbreed* 'hairbreadth' | 0.20 | 17; 5 | **1.***toegeven* 'to give in'(8), **2.***in* *de* *weg* *leggen*|*staan* 'to get|be in the way'(5), **3.***afwijken* 'to deviate'(2) |
| *kilometer* 'kilometer' | 0.0008 | 19; 8 | **1.***rijden* 'to drive'(10), **2.***kloppen* *voor* 'to be right', *omrijden* 'to make a detour'(2) |
| *meter* 'meter' | -0.02 | 100; 37 | **1.***werken* 'to work'(13), **2.***gaan* 'to go'(8), **3.** *kloppen* 'to be correct', *lopen* 'to run', *lukken* 'to manage', *zitten* 'to sit'(7) |
| *millimeter* 'millimeter' | 0.02 | 100; 41 | **1.***wijken* 'to give in; to give way'(14), **2.***toegeven* 'to give in'(13), **3.***afwijken* 'to deviate'(10) |
| *muis* 'mouse' | -0.06 | 19; 8 | **1.***bij* *kunnen* 'can be added'(7), **2.***doorheen* *gaan*|*komen* 'to go, get through'(5), **3.***in* *kunnen* 'to get in'(2) |
| *nacht* 'night' | -0.02 | 22; 5 | **1.***wakker* *liggen* 'to lie awake'(11), **2.***slapen* 'to sleep'(8), **3.***geheim* *blijven* 'to remain secret', *snurken* 'to snore', *uit* *zijn* *gedachten* *zijn* 'to be out of his mind, his thoughts'(1) |
| *schub* 'scale' | -0.004 | 13; 2 | **1.***vangen* 'to catch'(12), **2.***aan* *de* *haak* *slaan* 'to hook up'(1) |
| *spier* 'muscle' | -0.04 | 18; 4 | **1.***bewegen* 'to move'(10), **2.***verroeren* 'to stir'(6), **3.***helpen* 'to help', *veranderen* 'to change'(1) |
| *stap* 'step' | 0.03 | 77; 20 | **1.***vooruitkomen* 'to move forward'(39), **2.***lopen* 'to run'(6), **3.***opschieten* 'to make progress'(5) |
| *strobreed* 'strawbreadth' | 0.04 | 100; 3 | **1.***in* *de* *weg* *leggen* 'to get in the way'(97), **2.***toegeven* 'to give in'(2), **3.***afwijken* 'to deviate'(1) |

Table 13: Overview minimizers cluster 9: silhouette width, token frequency, type frequency, and top predicates

Despite the fact that the last cluster has a very low ASW (0.04), 8 of its members belong to the same semantic category: units of distance. Based on their distributional profile, we can distinguish three different groups: (i) *duimbreed*, *haarbreed*, *strobreed*, (ii) *centimeter*, *kilometer*, *meter*, *millimeter*, and (iii) *stap*.

i) First, *haarbreed* 'hairbreadth', *duimbreed* 'thumbsbreadth', and *strobreed* 'strawbreadth' refer to minimal distances, but, in contrast to the minimizers of group (ii), *haarbreed* and *strobreed* are not official measures. Moreover, in Dutch, *duim* is an obsolete measure of distance (cf. English *inch* and French *pouce*). As Table 13 shows, the three nouns of group (i) have a low type frequency, and they are often combined with the same type of concession verbs: *toegeven* 'to give in', *in de weg leggen* 'to get in the way', *wijken* 'to give in', and *afwijken* 'to deviate'. Note that the very idea of 'concession', i.e., *con-cedere*, is a natural metaphoric extension of motion verbs (to step aside, to retract oneself, etc.). This conceptual bridge located in the predicates themselves (cf. also cluster 2: *waard zijn* 'have a certain value; be capable'), facilitates the extension of the distance minimizers. *Strobreed* and to a lesser extent *duimbreed* have a clear preference for respectively *in de weg leggen* and *toegeven.*

ii) The more common measures of distance united in the second group (*centimeter*, *kilometer*, *meter*,and *millimeter*)are combined with a much wider array of verbs, as evidenced by their type frequency. *Kilometer* is the least frequently used one, which can be explained by the fact that it is less minimal than, for instance, *millimeter.* Most of the instances with *kilometer* clearly refer to distance, as illustrated by example (24a). However, in some examples (such as 24b and c), this distance meaning is eroded, possibly through the influence of the more common minimizer *meter* (cf. the parentheses in example 24b).

(24) a) *Ook al rijdt u geen kilometer, dan nog betaalt u uw verzekering en wegenbelasting*.

'Even if you don't drive a kilometer, you still pay your insurance and road tax.'

b) *Na een paar kilometer had ik wel in de gaten dat het een zware dag zou worden. Mijn knie werkte voor geen (kilo)meter mee*.

'After a few kilometers I realized it was going to be a tough day. My knee did not cooperate for a (kilo)meter.'

c) (i.v.m. voedingswaarden) *Lamssaucijs en lamsburger. Die klopte ook voor geen kilometer*.

(in the context of nutritional values) 'Lamb sausage and lamb burger. That one was also not right for a kilometer.'

The frequency spectrum below shows the wide variety of types *meter* is combined with, which explains its low – even negative – silhouette value (–0.02).

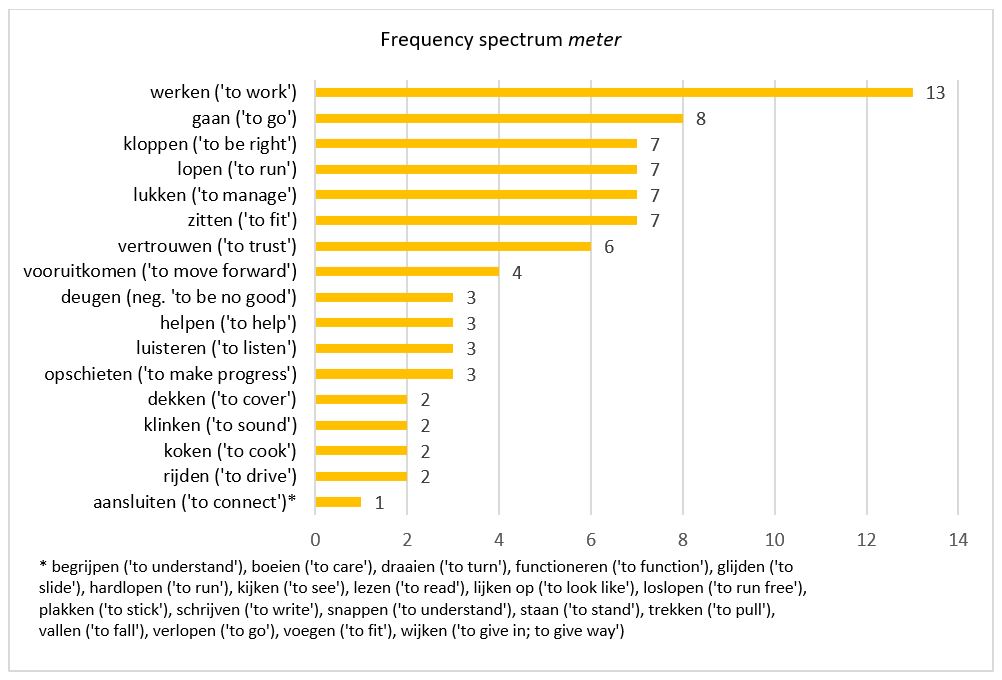


Figure 13: Frequency spectrum *meter*

Only in a few sentences, *meter* is still used to refer to distance (cf. examples 25a and 26a). In the majority of corpus instances, the combination of *meter* with the predicate metaphorically expresses a lack of progress, as in examples (25b) and (26b), and, by metonymical association, a lack of proper functioning (example 27) that causes a lack of progress[[16]](#footnote-16), which links this extension to the core predicates of this cluster.

(25) a) *Geen meter gelopen, geen seconde yoga en ook geen work-outs.*

'Didn’t walk a meter, not a second of yoga and no work-outs either.'

b) *Probleem: de communicatie loopt hier voor geen meter.*

'Problem: the communication (lit.) doesn’t run for no meter (= is going nowhere here).'

(26) a) *Je komt zonder speciale schoenen geen meter vooruit in de meters dikke sneeuw!*

'Without special shoes, you can't move a meter in the meters of thick snow!'

b) *Had wel internet op mijn telefoon, maar de verbinding is hier heel erg slecht dus dat schoot ook voor geen meter op.*

'Did have internet on my phone, but the connection here is really bad so that didn't make any progress either.'

(27) *De handschriftfunctie die wel interessant was - je schreef met een stylus en dat werd dan omgezet in getypt schrift - werkte eigenlijk voor geen meter.*

'The handwriting function that was interesting - you wrote with a stylus and this was then converted into typed script - did not really work for a meter.'

Next to the 'local' expansion of the micro-construction, based on the metaphorical and metonymical ties concerning progress verbs as described above, there is obviously also another source of expansion for distance minimizers, viz. analogical attraction by the prototypical minimizers of cluster 3 (the nearest cluster of *meter*). Thus, 13 of the 37 predicates of *meter* are also combined with taboo minimizers, among which the prototypical cognitive verbs *begrijpen* 'to understand', *geloven* 'to believe', and *snappen* 'to understand; to get it'. Both lineages lead toward a more internalized meaning of the verb-minimizer combination, which is a development typical of incipient subjectification (Traugott 1989), a type of semantic change linked to the grammaticalization process.

The other two minimizers of group ii), *millimeter* and *centimeter*,have a high type frequency and a similar distributional profile: they share no less than 16 different predicates. Their most frequent predicates, the concessive ones (*toegeven* 'to give in' and *wijken* 'to give in; to give way to'), are also combined with *haarbreed*, *strobreed*, and *duimbreed*, the first group. In combination with motion/progress verbs, however, they express more often physical motion (example 28a), compared to *meter*. The same metaphorical development can be witnessed, but to a lesser extent (examples 28b and 28c).

(28) a) *Voor mensen die een grote zeiltocht over zee aan het maken zijn is dat niet leuk. Je ligt stil in je boot en je komt geen centimeter vooruit.*

'For people who are taking a big sailing trip across the sea, this is no fun. You lie still in your boat and you don't move a centimeter.'

b) […] *dat wij met religie als zodanig geen centimeter vooruitkomen.*

'[…] that religion as such does not help us a centimeter.'

c) *beide partijen bewogen geen millimeter in de onderhandelingen.*

'neither side moved a millimeter in the negotiations.'

The fact that *centimeter* and *millimeter* are less subject to the drift undergone by *meter* can also be inferred from the limited presence of metonymically related predicates that denote 'proper functioning' (i.e., bad functioning = no progress) and cognition verbs. As to the latter, only *millimeter* is used once with *geloven* 'to believe'. However, there are some cases of host-class expansion in which *millimeter* and *centimeter* extend to verbs that are very infrequent at the level of the macro-construction:

(29) a) *Mijn letterhonger was weg, kon mij voor geen centimeter concentreren op het lezen.*

'My hunger for letters was gone, I could not concentrate on reading for a centimeter.'

b) *De publieke moraal werd en wordt vergiftigd door elite en establishment die zich […] geen millimeter schaamt om het onderste […] uit de kan te halen.*

'Public morality has been and is being poisoned by elites and establishment who are not a millimeter ashamed to make the most of it.'

Note that in example (29a), the predicate 'concentrate (during reading)' still has a metonymic link with the central sense 'lack of (physical) progress', viz. lack of concentration 'causes'lack of progress regarding the progressive scanning of the text.

iii) The third group of units of distance consists of only one minimizer: *stap* 'step'. *Stap* has some verbs in common with group i), viz. the concession verbs *afwijken*, *gaan*, and *wijken*, and shares a lot of predicates with group ii). The most frequent combination is *geen stap vooruitkomen* (lit. 'to make not a step progress'), which is used both in a spatial and metaphorical sense.

The other minimizers in this cluster, *muis*, *nacht*, *ding, spier*, and *schub*, are peripheral members with negative silhouette values. Of these five minimizers, *spier* 'muscle' shares three of its four predicates with the other nouns of this cluster: the progression predicates such as *bewegen* 'to move' also occur with *centimeter, millimeter*,and *meter.* The distributional profile of *spier* is also similar to that of *vin* (cluster 6), as both minimizers are combined with the semantically related verbs *bewegen* 'to move' and *verroeren* 'to stir; to move', referring to minimal motion or effort by means of the body parts involved in them.

*Nacht* 'night' only has one verb in common with the other nouns of the cluster: *slapen* 'to sleep' is combined once with *centimeter*. *Nacht* is actually more at home in its nearest cluster: the temporal minimizers, i.e., cluster 1. Admittedly, *nacht* is a quite specific time slot, as reflected in the most frequent predicates, which prevents *nacht* from having more shared predicates with the other temporal minimizers (only *slapen* 'to sleep' and *uit zijn gedachten zijn* 'to be out of his thoughts'). Yet, the number of shared predicates is larger than those shared with cluster 9.

*Schub*, *muis*,and *ding* have no predicates in common with other minimizers of this cluster, so their cluster assignment is actually almost random. *Ding* 'thing' is not at home in any cluster, being an infrequent minimizer, mostly used in biblical contexts, and it is combined with one single predicate that is unique in the entire dataset (*wees in geen ding bezorgd* 'don’t worry about a thing').

*Schub* 'scale', which metonymically (pars pro toto) refers to a fish, would, based on this meaning, be better at home in cluster 6. However, the only link between *schub* and the rest of the dataset is the fact that some nouns of cluster 3 (the nearest cluster of *schub*) are also combined with its predicate *vangen* 'to catch':

(30) a) *Zal de rest van het jaar wel geen biet meer vangen na deze uitspraken, maar dat is dan niet anders.*

'Probably won't catch a beet for the rest of the year after these statements, but that's that.'

b) […] *sportvissers die geen moer meer vangen* […]

'[…] anglers who no longer catch a MOER[*…*]'

*Muis* 'mouse' has a unique profile. Its nearest cluster, according to the algorithm, is cluster 5: *muis* and *drup(pel)* 'drop; drip' share three predicates (*bij kunnen* 'can be added', *doorkomen* 'to come through', and *doorlaten* 'to let through'). However, semantically, it is related to the animate nouns of cluster 6, with which *muis* also shares three predicates. Often, *muis* symbolizes an extremely small animate referent, which only needs a few spaces to slip through (as is the case for a drop, which explains the analogy):

(31) *Zoals jaarlijks het geval is, kan er daardoor geen muisje meer bij tijdens de feestavonden op zaterdag- en maandagavond.*

'As is the case every year, this means that there is no more room for a mouse (diminutive) during the party evenings on Saturday and Monday.'

*Muis* is also frequently used in the context of football (six out of 19 sentences). The defense simply leaves no room to slip through, for nobody, not even the smallest animal:

(32) *Stel je voor dat hij naast Kompany zou staan bij Manchester City, met Nigel de Jong ervoor. Daar komt, om in voetbaltermen te blijven, geen muisje meer doorheen.*

'Imagine him standing next to Kompany at Manchester City, with Nigel de Jong in front of him. In football terms, not a mouse (diminutive) could get through that.'

In sum, the fact that five of the 13 minimizers have negative silhouette values explains why cluster 9 has the lowest ASW. They have very unique distributional profiles with only a very small, or even no, link at all with other minimizing nouns in the dataset because they hardly share any predicate with other clusters. As a result, these nouns are by definition hard to cluster.

**4. Discussion and conclusion**

In this study we determined the semantic structuring of the different minimizing constructions in present-day Netherlandic Dutch on the basis of the predicates they combine with. Compared to previous corpus-based work by Hoeksema (2001; 2002) (from a diachronic perspective), our research here includes a larger set of minimizers, allowing for a more sophisticated quantitative and data-driven treatment of their distribution, followed by a detailed qualitative analysis, showing several flavors of context expansion.

The clusters generated by the algorithm have a clear semantic core. The predicates that are typically combined with certain semantic classes (or subclasses, e.g., in cluster 8: verbs of written vs. oral communication) of minimizers are still in harmony with the referential meaning of those minimizers. This suggests that, despite many extensions, a considerable number of minimizers preserved (at least a core of) unbleached usage.

The clusters were ranked from high to low average silhouette width, with the time minimizers being the most robust category. Some minimizers were hard to cluster because of their particular nature, which can be inferred from their low silhouette values. A first case in point are minimizing nouns that are part of nearly fixed expressions, such as *vlieg* 'fly' (cluster 6). These exhibit extremely high proportions for one predicate that hardly occurs with other minimizers.Another example of such particular behavior is found in'singletons' such as *schub* 'scale' (cluster 9) and *korrel* 'granule; grain' (cluster 6), with almost no ties to the rest of the dataset. In these cases, clustering assignment is nearly random. A third kind of complication for clustering concerns minimizers that appear to be blends of two clusters. This results in marginal class membership and competition with another equally marginal class assignment (i.e., the nearest cluster proposed by the algorithm). For instance, *drol* 'turd' is a mix of cluster 3, and cluster 2, *jota* 'iota' is situated in cluster 4 but close to cluster 3, and *kruimel* 'crumb' belongs to cluster 7 but is close to cluster 5. Note that the hesitation of the algorithm is in itself meaningful: for instance, a turd being a scatological object (cluster 3) with no value (cluster 2).Finally, a traditional drawback of type-based distributional methods is that they ignore polysemy (cf. *vinger* 'finger', cluster 7).

More interestingly, our method, and especially low silhouette values, allowed us to detect not only different types of lexical and semantic host-class expansion but also analogical links within the paradigm of minimizing constructions. In some cases, the expansion seems to be situated merely at the level of a single micro-construction, for instance, if this construction is combined with predicates that are hapaxes in the entire dataset (e.g., *kwartje* 'quarter' [cluster 2], *millimeter* 'millimeter', and *centimeter* 'centimeter' [cluster 9]). An extension can also involve an analogy between several micro-constructions that form part of the same cluster. Such local extensions can be explained by semantic processes such as metaphorization and generalization, affecting the predicates (e.g., cluster 9: progress, concession) or the minimizers themselves (e.g., cluster 2: *brood* 'bread', monetary value ® worth; cf. also *pepernoot* 'ginger nut'), yielding more internal senses, an evolution which is often associated with incipient subjectification (Traugott 1989). We also detected cases of extension that seem to be based on subtle metonymic associations, conceptual bridges based on some contiguity, through which unexpected predicates can be linked with more central predicates, such as bad functioning causing lack of progress (*meter* 'meter'), success generating (monetary) value (*pepernoot* 'ginger nut', in itself already metaphorically used), or, combined with metaphors, trust being considered as a value to be exchanged within a fictive monetary transaction (*voor geen cent vertrouwen* 'not to trust for a cent'). All these processes yield lexical and semantic host-class expansion (Himmelmann 2004), which appears as a natural development of the scope of the micro-constructions (or their clusters) themselves.

Besides these local extensions, we saw micro-constructions taking predicates that are rather unexpected on the basis of their original referential meaning. On closer inspection, these predicates correspond to the most frequently occurring predicates in the macro-construction considered as a whole, especially cognition verbs and verbs of indifference (e.g., *pepernoot* 'ginger nut' [cluster 2], *jota* 'iota' [cluster 4], and *meter* 'meter' [cluster 9]). These predicates are also the hallmark of the taboo minimizers, the biggest cluster, and hence the principal provider of predicates in the dataset. Their salience seems also to be boosted by the analogical attraction among taboo minimizers, which indeed often share the same limited set of predicates. Their status as 'purveyor to the court' also appears in the number of tokens they provide, since many taboo minimizers attain the maximum of 100 tokens. Their prominent status is also reflected in their overall (extrapolated) token frequency; many of them are among the most frequently occurring minimizers in Netherlandic Dutch. Speakers are hence often exposed to them in discourse. For all these reasons, they can definitely be viewed as the most prototypical minimizing nouns in Dutch. In other words, it seems as if the macro-construction as a whole, via its dominating predicates, exerts a strong pressure on individual micro-constructions to extend in a direction that is, at least from the point of view of the individual micro-constructions, semantically non-harmonious, contributing to their further relaxing of semantic restrictions and bleaching.

These distinct flavors of context expansion hint at the grammaticalization process of minimizers as negation-reinforcing devices (Hansen & Visconti 2009). Admittedly, they illustrate only but incipient grammaticalization (Breitbarth *et al*. 2013); none of these minimizing constructions seems on its way to becoming a new negator. The mechanisms at hand are related to lexical and semantic host-class expansion. They all illustrate processes of semantic extension/generalization, leading quite often to the addition of more internal senses, which may be seen as a new step toward subjectification, on top of the default 'subjective' stance implied in the hyperbolic use (by means of scale reversal) of the nouns involved, which, indeed, even when they seem to have maintained their referential meaning, already act as a conventionalized symbol for 'nothing/not at all'. E.g., if one says, 'I have not missed a word of this paper', then this is already a subjective, hyperbolic expression, since it is very unlikely that no single word has been skipped. Considerably more advanced features of subjectification are found with taboo terms (and associates such as *hout*, etc.), which are extravagant negation intensifying devices with no link to semantic-referential content whatsoever. Surprisingly, these purportedly semantically 'empty' nouns are collocationally constrained as they combine with a rather limited set of strongly overlapping predicates.

In sum, our analysis reflects interesting mechanisms of ongoing context expansion that can be related to incipient grammaticalization and subjectification. It should be noted, however, that these hypotheses are based on a synchronic comparison of distributional profiles. They should be confirmed by a genuinely diachronic perspective.

**References**

Barðdal, J. (2008). *Productivity: Evidence from case and argument structure in Icelandic* (= Constructional approaches to language 8). Amsterdam: John Benjamins.

Breitbarth, A., C. Lucas & D. Willis. (2013). Incipient Jespersen’s Cycle: the (non-)grammaticalization of new negative markers. In: J. Fleischer & H. Simon (eds.), *Comparing Diachronies*. Berlin/Boston: De Gruyter, 141–162.

Divjak, D. & N. Fieller (2014). Cluster analysis. Finding structure in linguistic data. In: D. Glynn & J. A. Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*. Amsterdam: John Benjamins, 405–441.

Divjak, D. & S. T. Gries (2006). Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2(1), 23–60.

Firth, J.R. (1957). *Papers in Linguistics, 1934-1951*. London: Oxford University Press.

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure.* Chicago: University of Chicago Press.

Gries, S. T. & A. Stefanowitsch (2010). Cluster analysis and the identification of collexeme classes. In: S. Rice & J. Newman (eds.), *Empirical and experimental methods in cognitive / functional research*. Stanford, CA: CSLI. 73–90.

Gyselinck, E. & T. Colleman (2018). *The role of expressivity and productivity in (re)shaping the constructional network: a corpus-based study into synchronic and diachronic variation in the intensifying fake reflexive resultative construction in 19th to 21st Century Dutch. Proefschrift.* Ghent: Ghent University.

Hansen, M. B. Mosegaard & J. Visconti. (2009). On the diachrony of “reinforced” negation in French and Italian. In: C. Rossari, C. Ricci & A. Spiridon (eds.), *Grammaticalization and pragmatics: Facts, approaches, theoretical issues*. Bingley (UK): Emerald, 137–171.

Hansen, M.B. Mosegaard. (2013). Negation in the history of French. In D. Willis, C. Lucas & A. Breitbarth (eds.), *The history of negation in the languages of Europe and the Mediterranean. Volume 1: Case studies.* Oxford, UK: Oxford University Press, 51–76.

Haspelmath, M. (1999). Why is grammaticalization irreversible? *Linguistics* 37(6), 1043-1068.

Heylen, K., T. Wielfaert, D. Speelman & D. Geeraerts (2015). Monitoring polysemy: Word space models as a tool for large-scale lexical semantic analysis. *Lingua* 157, 153–172.

Himmelmann, N. P. (2004). Lexicalization and grammaticization: Opposite or orthogonal. In: W. Bisang, N. P. Himmelmann & B. Wiemer (eds.), *What makes grammaticalization: A look from its fringes and its components*. Berlin: Mouton de Gruyter, 21–42.

Hoeksema, J. (2001). Rapid change among expletive polarity items. In: L. J. Brinton (ed.), *Historical linguistics 1999. Selected papers from the 14th international conference on historical linguistics, Vancouver, 9-13 August 1999*. Amsterdam: John Benjamins, 175–186.

Hoeksema, J. (2002). Minimaliseerders in het Standaardnederlands. *Tabu* 32, 105–174.

Hoeksema, J. (2003). Partitivity, degrees and polarity. *Verbum* XXV, 81–96.

Hoeksema, J. (2019). Taboo terms and their grammar. In: K. Allen (ed.), *The Oxford Handbook of Taboo Words and Language*. Oxford: Oxford University Press, 160–179.

Kaufman, L. & P. J. Rousseeuw (1987). Clustering by means of medoids. In: Y. Dodge (ed.), *Statistical data analysis based on the L1-norm and related methods*. Amsterdam: North-Holland, 405–416.

Kaufman, L. & P. J. Rousseeuw (2005). *Finding groups in data: an introduction to cluster analysis*. Hoboken (N.J.): Wiley.

Keller, R. (1994). *Sprachwandel: von der unsichtbaren Hand in der Sprache. 2., uberarb. und erw. Aufl*. Tübingen: Francke.

Lakoff, G. (1987). *Women, fire, and dangerous things: what categories reveal about the mind*. Chicago: University of Chicago press.

Perek, F. (2016). Using distributional semantics to study syntactic productivity in diachrony: a case study. *Linguistics* 54(2), 149–188.

Perek, F. (2018). Recent change in the productivity and schematicity of the *way*-construction: a distributional semantic analysis. *Corpus Linguistics and Linguistic Theory* 14(1). 65–97.

Perek, F. & M. Hilpert (2017). A distributional semantic approach to the periodization of change in the productivity of constructions. *International Journal of Corpus Linguistics* 22, 490–520.

Postma, G. (2001). Negative polarity and the syntax of taboo. In: J. Hoeksema, H. Rullmann, V. Sanchez-Valencia, T. van der Wouden (eds.), *Perspectives on negation and polarity items*. Amsterdam: John Benjamins, 283–330.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20, 53–65.

Stefanowitsch, A. & S. T. Gries. (2003). Collostructions: Investigating the interaction of words and constructions. *International journal of corpus linguistics* 8(2), 209–243.

Traugott, E.C. (1989). On the rise of epistemic meanings in English: an example of subjectification in semantic change. *Language* 65(1), 31–55.

Traugott, E. C. & G. Trousdale (2013). *Constructionalization and constructional changes*. Oxford: Oxford University Press.

Turney, P. & P. Pantel (2010). From frequency to meaning: vector space models of semantics. *Journal for Artificial Intelligence* 37, 141–188.

Usino, W., A. S. Prabuwono, K. H. S. Allehaibi, A. Bramantoro, A. Hasniaty & W. Amaldi (2019). Document similarity detection using k-means and cosine distance. *International journal of advanced computer science and applications* 10, 165–170.

Van den Heede, M. & P. Lauwers (2021). Het interesseert me geen ene je-weet-wel. Minimaliserende strategieën in het hedendaagse Nederlandse Nederlands. Handelingen van de Koninklijke Zuid-Nederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis 75(1), 239–259.

Van den Heede, M. & P. Lauwers (2023). Syntactic productivity under the microscope: The lexical and semantic openness of Dutch minimizing constructions. *Folia Linguistica* 57(3), 723–761.

Vandeweghe, W. (2005). Minimaliseerders en negatief gebonden *of*-constructies. *Voortgang, Jaarboek voor de Neerlandistiek* 23, 239–247.

*\* COW corpus*:

Schäfer, R. (2015). Processing and querying large web corpora with the COW14 architecture. In: *Proceedings of challenges in the management of large corpora* (CMLC-3) (IDS publication server), 28–34.

Schäfer, R. & F. Bildhauer (2012). Building large corpora from the web using a new efficient tool chain. In: *Proceedings of the eighth international conference on language resources and evaluation* (LREC’12), 486–493.

texrex <https://github.com/rsling/texrex>

COW tool chain <https://github.com/rsling/cow>

\* *R and RStudio*:

R Core Team (2023). R: A language and environment for statistical computing. *R foundation for statistical computing*. Vienna, Austria. <https://www.R-project.org/>

Posit team (2024). RStudio: Integrated development environment for R. *Posit Software, PBC*. Boston, MA. <http://www.posit.co/>

\* *R packages*:

Hennig, C. (2020). fpc: Flexible Procedures for Clustering. R package version 2.2-9.

Kassambara, A. & F. Mundt. (2020). factoextra: Extract and visualize the results of Multivariate Data Analyses. R package version 1.0.7.

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert & K. Hornik. (2019). cluster: Cluster Analysis basics and extensions. R package version 2.1.0.

Plevoets, K. (2015). *svs: Tools for semantic vector spaces*. <https://cran.r-project.org/web/packages/svs/>

\* *TenTen corpus*:

Jakubíček, M., A. Kilgarriff, V. Kovář, P. Rychlý & V. Suchomel (2013). The TenTen corpus family. In: *7th International Corpus Linguistics Conference CL*. Lancaster, 125–127.

\* *Websites*:

Bible Hub:

<https://biblehub.com/matthew/5-18.htm>

Stack Exchange:

<https://stats.stackexchange.com/questions/120350/k-means-on-cosine-similarities-vs-

euclidean-distance-lsa>

TROLLing data repository:

<https://dataverse.no/dataverse/trolling>

Vlaanderen, Team Taaladvies:

<https://www.vlaanderen.be/taaladvies/taaladviezen/letterlijk-betekenis>

Woordenboek der Nederlandse taal:

<https://ivdnt.org/woordenboeken/woordenboek-der-nederlandsche-taal/>

1. The annotated dataset will be stored on the TROLLing repository <https://dataverse.no/dataverse/trolling>. [↑](#footnote-ref-1)
2. Based on the ratio of true positives/tokens viewed, projected against the initial number of hits. [↑](#footnote-ref-2)
3. The frequencies are unweighted. Weighting by means of a positive pointwise mutual information score (PPMI) lowered the average cluster quality (average silhouette width [ASW]) to 0.08 probably because of the fact that PPMI overestimates the many low frequent combinations in this dataset. [↑](#footnote-ref-3)
4. Although k-means is sometimes used with cosine similarity (e.g., Usino *et al*. 2019), it remains a point of debate. See, for instance, <https://stats.stackexchange.com/questions/120350/k-means-on-cosine-similarities-vs-euclidean-distance-lsa> (last accessed 30th of May 2022). [↑](#footnote-ref-4)
5. One of our reviewers asked why we used ASW in order to determine the optimal number of clusters rather than the elbow method or the gap statistic. We used average silhouette both for its precision (compared to the elbow method) and its simplicity (no bootstrapping, contrary to the gap statistic). This choice is also consistent with the use of silhouette widths throughout the paper to detect outliers. [↑](#footnote-ref-5)
6. Note that this value is rather low and hints at a 'weak' structure that 'could be artificial' (Kaufman & Rousseeuw 2005: 88). However, for this kind of clustering based on dissimilarity matrices, we experienced that silhouette values in this range – and even worse – still correspond to highly meaningful clusters. For instance, we tested the PAM algorithm and its silhouette widths with three classes of 10 verbs in Dutch (verbs of motion, communication, and disintegration). Clustering of their dissimilarities (stemming from word embeddings generated with the Dutch COW corpus, Schäfer 2015; Schäfer & Bildhauer 2012) was able to detect the 3 initial groups almost perfectly, although the silhouette widths were still below 0.25. [↑](#footnote-ref-6)
7. Eight *duit* pieces are worth one *stuiver* (cf. WNT: <https://ivdnt.org/woordenboeken/woordenboek-der-nederlandsche-taal/>, last accessed 20th of June 2023). [↑](#footnote-ref-7)
8. Note that we conceive this value here as a rough measure of semantic variability of (the predicates of) each class rather than as a genuine productivity measure like type/token ratio (which, anyway, would be incomparable because of unequal sample sizes of the classes). We refer the reader to Van den Heede & Lauwers (2023), for a detailed comparison of the productivity of the micro-constructions found with at least 100 tokens. [↑](#footnote-ref-8)
9. Sometimes, we list more than five predicates if some of the predicates are part of nearly fixed expressions (for instance, cluster 5), which has a major influence on their frequencies. [↑](#footnote-ref-9)
10. Recall that the algorithm seeks the best global solution. [↑](#footnote-ref-10)
11. All examples are from the nlTenTen14 corpus (cf. Jakubíček *et al*. 2013). [↑](#footnote-ref-11)
12. Our dataset contains several examples of this type of creative language use, such as: *Nog een vraag waar letterlijk geen drol van klopt is de vraag over hondenpoep* 'Another question that literally doesn't make sense a turd is the one about dog shit.' As pointed out by one of our reviewers, *letterlijk* has two different functions: a metadiscursive gloss pointing to literal language use and an intensifier of figurative language (e.g., *De patiënt is letterlijk door het oog van de naald gekropen* 'The patient literally crawled through the eye of the needle,' <https://www.vlaanderen.be/taaladvies/taaladviezen/letterlijk-betekenis>, last accessed 20th of June 2023). In our corpus, both functions seem to conspire: *letterlijk* acts as an additional intensifier of the minimizer, while at the same time, it recalls the link with the referential context (dogs ® dog shit) that justifies the creative extension of the minimizer. [↑](#footnote-ref-12)
13. There are different types of taboo minimizers: religious taboo terms (e.g., *bliksem* 'lightning'), sexual and scatological taboo terms (e.g., *fluit* 'flute', *drol* 'turd'), and taboo terms denoting contagious or lethal diseases (e.g., *pest* 'plague') (Hoeksema 2001: 3–7). [↑](#footnote-ref-13)
14. Also register-related differences may be involved in the collocational preferences exhibited by minimizers (Hoeksema 2002: 42–51). As our corpus does not allow us to quantify such register-based contrasts, we won’t develop this point further. [↑](#footnote-ref-14)
15. By analogy with the use of *jota* in the Bible (*Matth.* 5, 18): 'For truly, I say to you, until heaven and earth pass away, not an iota, not a dot, will pass from the Law until all is accomplished' (<https://biblehub.com/matthew/5-18.htm>, last accessed 20th of June 2023). [↑](#footnote-ref-15)
16. Note that there is also an import syntactic aspect in the expansion of *meter*, namely the frequent combination with the preposition *voor* 'for'. Not only in the case of *meter*, but also for other micro-constructions, the preposition has an impact on their extension (cf. examples 8, 9, and 29a). However, a detailed analysis of the role of the preposition goes beyond the scope of this paper. [↑](#footnote-ref-16)