Words that second language learners are likely to hear, read, and use*


In the present study, we explore whether multiple data sources may be more effective than single sources at predicting the words that language learners are likely to know. Second language researchers have hypothesized that there is a relationship between word frequency and the likelihood that words will be encountered or used by second language learners, but it is not yet clear how this relationship should be effectively measured. An analysis of word frequency measures showed that spoken language frequency alone may predict the occurrence of words in learner textbooks, but that multiple corpora as well as textbook status can improve predictions of learner usage.

Research on vocabulary knowledge in second language (L2) learners has for the most part concentrated on developing measures to assess the proficiency or vocabulary size of individual learners. This study addresses a different but related question: how does an experimenter or researcher choose words that learners of an L2 are likely to know during acquisition? Experimenters may need to select words likely to be known by learners to test non-lexical aspects of L2 knowledge. Similarly, when constructing materials for students, instructors must in some cases choose materials so that learners are likely to already know the words in a lesson, so that other, non-lexical aspects of the lesson are conveyed effectively. For child language learners there are vocabulary tests such as the Peabody Vocabulary Test (Dunn and Dunn, 1997) available with estimates regarding which words are known at a particular age. No such tests exist for second language learners. Conversely, researchers will sometimes require words that learners are unlikely to know, if a study is designed to examine how learners process or learn unfamiliar words. On what basis should these choices be made?

A straightforward answer seems to be that words should be selected on the basis of their lexical frequency. Word frequency is an important factor in second language acquisition (SLA), as learners tend to recall (Laure, Elder, Hill and Congdon, 2004) and recognize (Brown, 1993) words that are more common in a given language better than less frequent words. Based on this relationship between lexical statistics and vocabulary knowledge, it is generally accepted that knowing less frequent words is indicative of having a larger vocabulary or, as Meara (2005, p. 32) puts it, “[the] use of frequency bands to characterize vocabulary is a fairly standard practice in L2 vocabulary studies”. However, the available word frequency data are based on samples from different sources such as newspapers, books and conversation, and this leads to the problem of whether estimates from all sources are equally appropriate for the population of second language learners. Samples of language use from which a set of words is selected may be biased in the sense that some of the properties of the sample words may not represent accurately the properties of the entire language from which it is sampled while other properties may in fact be represented accurately. In newspaper text, for example, words related to certain topics, such as politics, are typically overrepresented. When constructing materials for experiments or teaching purposes, the bias of the corpora should not enter as a bias in the materials. More importantly, the language to which learners are exposed and therefore the vocabulary they acquire, is itself biased. Even an ideal sample that represents the entire language correctly would not necessarily yield a frequency estimate that optimally predicts the vocabulary of learners. Most clearly this is the case for classroom learning where the frequency of words that learners

* Arna van Doorn assembled the vocabulary lists from the three Dutch textbooks. The Max Planck Institute for Psycholinguistics provided access to the CELEX, the CGN, and the ESF corpora. The analysis was conducted using R (R Development Team, 2005), and the stats (R Development Team, 2005) and MASS (Venables & Ripley, 2002) libraries. This research was supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). We would also like to thank two anonymous reviewers for useful suggestions, and Jan Hulstijn for providing helpful comments and references for textbook vocabulary selection including, in addition to those cited in the text, Hazenberg (1994) and Sciarone (1979).

Address for correspondence:
Doug Davidson, F. C. Donders Centre for Cognitive Neuroimaging, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands
doug.davidson@fc-donders.ru.nl
encounter does not reflect the entire language, but rather the teaching materials. It is known that words that are more commonly encountered in school are more likely to be known on vocabulary tests. This has been demonstrated in experimental studies of vocabulary acquisition in reading (Pitts, White and Krashen, 1989; Day, Omura and Hiramatsu, 1991; Dupuy and Krashen, 1993; Hulstijn, Hollander and Greidanus, 1996; Rott, 1999), as well as through an empirical demonstration that the frequency of words presented in classroom materials predicts the likelihood of a correct answer on tests of vocabulary (Vermeer, 2001), and demonstrations that textbook appearance can be used as a predictor for word recognition ability (Van Gelderen et al., 2004; Fukkink, Hulstijn and Simis, 2005).

The problem of the classroom bias suggests the possibility that the occurrence of words in teaching materials such as textbooks may in fact be more informative with respect to early vocabulary than corpus frequencies, at least for classroom learners. Biases of the language that untutored learners are exposed to are obviously much more variable depending on the context of language input. Assuming, however, that textbook authors have a good intuition about the kind of vocabulary all language learners, irrespective of their learning circumstances, need, it is possible that textbook occurrence may be a good predictor for the early vocabulary of untutored early learners as well.

The goal of the present paper is to assess whether multiple data sources may be more effective than single data sources to predict the words that language learners are likely to know. We will first address the question of how word frequencies estimated from corpora are related to textbook vocabulary. Do frequency estimates from text corpora predict textbook occurrence with equal effectiveness as those from spoken corpora? In a second step, we will describe an estimation procedure using several data sources at once. We will assess which combination of word frequency data from speech and text corpora and occurrence in learner textbooks best predicts the occurrence of nouns in a corpus of untutored L2 learners’ speech. The goal is to find nouns that learners are likely to know across a broad range of settings both inside and outside the classroom.

Method
In the analyses reported below, we investigate the relationship between three types of information: (1) Estimates of noun form frequency taken from spoken and written corpora representing a wide variety of word usage, (2) vocabulary lists taken from adult foreign language learning textbooks, and (3) estimates of word usage taken from a small sample of untutored language learner speech based on a film description task. The goal for the analysis is to characterize the relationship between different corpus frequencies and to evaluate how well different corpus frequencies, learner textbook occurrence, and different combinations of these data sources predict actual learner production. Based on the best-predicting combination, the analysis should provide a list of words that second language learners are likely to know and use.

Word frequency lists typically contain a list of word forms paired with the number of instances of the word form in a corpus sample. Word frequency distributions are heavily skewed with the majority of word types having very low frequencies. This skew is a problem for researchers who wish to select common words because the numerical value of any frequency estimate will depend heavily on the size of the sample of speech or text. This makes it difficult to choose a fixed frequency count as a threshold for common words. In the example lists described below, we selected from word form frequency lists words with frequencies above the log average frequency based on words with frequencies > 1 in order to reduce the influence of very low frequency words. Rank information could be used in place of frequency counts, but frequencies potentially contain information that is lost in ranks. There may be a large difference between two rank-adjacent words in frequency, yet another pair of rank-adjacent words may have a small difference in frequency.

Two corpora were used for this study, the CGN (Corpus Gesproken Nederlands or the Spoken Dutch Corpus) and CELEX. The CGN word frequencies (Oostdijk, 2000; based on version 1.0 of the Spoken Dutch Corpus) were based on nouns taken from the CGN Lexicon portion of the corpus (Dutch and Flemish, comprising 181,579 word forms). Frequencies were scaled to frequencies per million. Only the single word-form frequencies were included in the analysis presented (the corpus includes information for multiple-word forms (phrases) in a separate lexicon). The corpus consists of samples of conversational and non-conversational speech including telephone conversations, debates, broadcast news reports and commentary, speeches, and lectures (a detailed description of the components is available in the documentation of the distributed corpus).

The CELEX word frequencies (Baayen, Piepenbrock and Gulikers, 1995) were based on the word form and word lemma frequencies included in the CELEX database. CELEX frequencies for Dutch are based on a sample of fiction (30%) and non-fiction (70%) books published between 1970 and 1988. The word frequency estimates do not include material from newspapers, magazines, children’s books, or textbooks.

The words from the textbooks were taken from three modern introductory language textbooks for Dutch (selected because they were conveniently available): Taal vitaal: Nederlands voor beginners (Schneider-Broekmans, 2000), Dutch for self-study (van Kampen
and Stumpel, 2002), and Colloquial Dutch: The complete course for beginners (Donaldson, 1996). All three textbooks are aimed at beginning learners of Dutch and include word lists. For the word form-based analyses described below we expanded the set of words from the textbook word lists (which only provide citation forms) by adding to it all noun forms of the lemma. If the word *kat* “cat” appeared in the textbook list, for example, then *katten* “cats” and *katje* “cat, diminutive”, and *katjes* “cats, diminutive” were added to the expanded textbook list.

Textbook wordlists, including the textbooks chosen in the present study, are likely to be selected based on word frequency (Bossers, 1996). Beginner textbooks concerning Dutch commonly use a wordlist compiled in De Kleijn and Nieuwberg (1983), based on frequency estimates produced in Uit den Bogaart (1975). While it would therefore be expected that word frequency would predict whether a word will be found on a textbook wordlist, earlier word frequency estimates have largely been based on text, and not speech, as large speech corpora have only become available recently. It is therefore not clear whether both types of frequency estimates would contribute to the prediction. In the present paper, the relative predictive power of text versus speech word frequency estimates is compared.

The learner production data consisted of words taken from a film-re-telling task collected as part of the European Science Foundation (ESF) Second Language Learner corpus (see Perdue, 1984; 1993 for details). The film re-tellings covered restricted portions of Chaplin’s film “Modern Times”. The corpus consisted of the film descriptions of four Moroccan and four Turkish immigrant learners of Dutch on two measurement occasions (two re-tellings of the same film) within the first year of learning Dutch. As such, the film re-tellings represent a sample of L2 production in which the same participants produced descriptions of a single film, in many ways like an experimental setup in which different participants produce sentences to an experimental stimulus set. The learners in this sample were untutored, that is, learners without exposure to classroom teaching and textbook materials.

For the nouns that were found in the Dutch learner textbooks, we estimated how well the spoken and written Dutch corpora (CGN and CELEX) frequencies predicted the status of a word as a textbook vocabulary entry (yes or no) using logistic regression. We use logistic regression to determine how strongly spoken and textual frequencies are related to the propensity to be found in an introductory Dutch textbook wordlist. For the analyses presented below, we compared model fits using deviance (p-values are reported for 1-df Chi-square comparisons), as well as the Akaike Information Criterion (AIC; Akaike, 1974) using stepwise comparison (both forward and backward) of model terms. AIC is a criterion for selecting between different statistical models (such as regression models), and is used widely in regression analysis. It is defined as 

\[-2 \log L -2m\]

where \(L\) is the log likelihood of the model under consideration, and \(m\) is the number of parameters in that model. The general benefit of applying AIC is that it takes into account both the goodness of fit (the likelihood of the model) as well as the number of parameters in the fitted model (the model complexity); smaller AIC values indicate a better-fitting model. Normalized differences in AIC values between competing fits can be interpreted as proportional likelihoods. To ensure that the results of our analyses did not depend on assumptions about the exact nature of the entries in the learners’ mental lexicon (e.g. whether or not unanalyzed inflected forms are listed), we estimated the fit based on word forms as well as on lemmas (summed frequencies of all forms of a word).

**Results**

**Corpus predictions of textbook vocabulary status**

Figure 1 shows the empirical proportion of word forms present in the textbooks and the predicted proportions for the models including CGN or CELEX for the range of scaled log CGN and CELEX form frequencies. This plot shows how the more common words in either of the corpora are more likely to be found in the textbooks, especially for the range of standardized frequency from 1.0 to 3.0. A model that included a positive term for CGN log frequency was the best-fit model (AIC = 4541), compared to CELEX alone (AIC = 5039). The model that included both CELEX and CGN (AIC = 4540) did not significantly improve the fit over the CGN-alone model \((p = 0.1)\). The parameters for the best-fitting model were

\[\text{Proportion}_\text{Textbook} = -0.7039 + 1.1959 \times \text{Frequency}_\text{CGN}\]

An additional analysis was conducted by using CELEX and CGN lemma frequencies rather than form frequencies. Figure 2 shows the empirical proportions of lemmas present in the textbooks and the predicted proportions based on the models including CGN or CELEX for the range of scaled log CGN and CELEX lemma frequencies. Note that given that there are more low frequency than high frequency words, the parameters of the best-fitting models will be such that the predictions for the low frequency ranges are better. It then appears that predictions based on CELEX underestimate textbook appearance for higher frequency items. Results of this analysis were similar to the previous analysis. The best-fitting model included only CGN log lemma frequency (AIC = 3057), compared to CELEX alone (AIC = 3390; \(p < 0.001\)), and the conjunction of CGN and CELEX frequencies (AIC = 3062) did not improve the fit \((p = 0.48)\) over CGN alone. The parameters for the best-fitting
Figure 1. Status as a textbook vocabulary item (0 = no, 1 = yes) based on form frequency. Solid line is the predicted proportion based on the fitted model with CGN as a predictor, dashed line with CELEX; open circles are the empirical proportions for CGN calculated at histogram intervals of the centered and scaled frequency distribution, filled circles are the CELEX proportions.

Figure 2. Status as a textbook vocabulary item (0 = no, 1 = yes) based on lemma frequency. Solid line is the predicted proportion based on the fitted model with CGN as a predictor, dashed line with CELEX; open circles are the empirical proportions for CGN calculated at histogram intervals of the centered and scaled frequency distribution, filled circles are the CELEX proportions.
model were Proportion_Textbook = \(-1.1071 + 1.3528 \times \) Frequency_CGN.

Appendix A contains a list of the 200 noun lemmas with the highest predicted probability of being found in a textbook based on this equation. For example, for the lemma *tijd* “time”, the predicted proportion of 0.9833 is obtained by log transforming the CGN frequency (i.e., LOG(1 + 903)) = 6.806829, scaling this value according to the mean (2.224635) and standard deviation (1.195641) of the log CGN frequencies (i.e., \(6.806829 - 2.224635/1.195641 = 3.832417\)), multiplying this value by the model coefficients (i.e., \(-1.1071 + 1.3528 \times 3.832417 = 4.0774\)) and substituting this value into the equation for the logistic (i.e., \(1/(1 + \exp(-4.0774)) = 0.9833\)).

In sum, spoken word form frequencies predicted the occurrence of words in learner textbooks better than written word form frequencies, and taking written word form frequencies into account did not improve the prediction based on spoken word forms alone. One interpretation of the results is that while both CELEX and CGN are correlated \((r = 0.765\) for log word form frequencies), there is a difference in variability of the frequency estimates. Apparently, some of the variability of the spoken word form frequencies predicts textbook status, but is not related to the variability in frequency that is shared with the written word form frequencies.

A possible explanation for the better fit of the model that is based on CGN frequency is that the vocabulary lists from the textbooks may include items that are common in colloquial speech, which may not be common in written text for stylistic reasons. Written text is often edited, for example. This could be due to characteristic terms used in speech and not in books, or vice versa. However, not all differences are entirely straightforward or predictable from genre differences: e.g. as seen in Appendix A, the lemma frequency for *soort* “sort”, is 1050 in CGN but only 75 in CELEX. Conversely, for *hand* “hand”, the frequency is 1028 in CELEX and 342 in CGN. This difference in frequencies would not be expected based on genre alone. This illustrates that different data sources may capture aspects of frequency estimates that are not otherwise easily modeled.

**Corpus and textbook predictions of learner usage status**

The previous analysis showed that the CGN corpus frequency better predicted whether or not a word appeared in textbook vocabulary lists than CELEX corpus frequency. It can, therefore, be assumed that spoken frequency predicts classroom vocabulary acquisition better than written frequency, given that the appearance of words in teaching materials predicts their likeliness of being learned in tutored SLA. We will now examine to what degree spoken or written corpus frequencies, as well as status as a textbook vocabulary item, predict whether a word is in the active vocabulary of early untutored learners of Dutch.

It can be expected that for a given sample of words that learners use, a large-corpus frequency estimate would predict whether or not a word appears in the sample because the more likely a word is in a language, the more likely it will be observed, by definition. What is important in the present case is whether or not one frequency estimate is better than another, or whether an ensemble of frequency estimates is better than any single estimate.

In contrast to the situation in classroom learning, which is highly structured according to a syllabus or a textbook, untutored language usage by learners is opportunistic, and structured according to the communicative needs of the learner. It is therefore possible that textbook status will not predict learners’ usage, along with, or in addition to, corpus frequency estimates. To investigate this question, we matched the nouns from the textbook list and corpora samples against the nouns appearing in a sample of spoken usage words from the film re-telling task of the Dutch portion of the ESF Second Language Learner corpus (Perdue, 1984, 1993).

Figure 3 shows the fit of the best-fitting model against the empirical proportions of words found in the film re-telling task as a function of form frequency for words found and not found in the textbooks. The best-fitting model included positive terms for both CGN and CELEX word frequency as well as textbook status (AIC = 1298); Proportion_Learner_Usage = \(-3.9373 + 0.6620 \times\) Textbook + 0.6793 \* CGN_Freq + 0.3550 \* CELEX_Freq. It should be emphasized that all three data sources were predictive of learner usage. However, in addition, the parameters of this model indicate that the change in proportion of learner usage related to Textbook status and CGN frequency was almost twice as large as the proportional increase related to CELEX frequency. The model with all three data sources was significantly better than models including textbook alone (AIC = 1503; \(p < 0.001\)), or textbook with CELEX alone (AIC = 1330; \(p < 0.001\)), or textbook with CGN alone (AIC = 1310; \(p < 0.001\)). Appendix B contains the forms ranked by the predicted proportion of the best-fitting model.

For a comparison to the previous analysis of frequency as a predictor for textbook status, we also fit models with CGN and CELEX without textbook status as a predictor for learner usage. The model parameters for the single term models (either CGN or CELEX) were of similar sign and magnitude as in the previous analysis. A model that included positive terms for CGN as well as CELEX log frequency was the best fit model (Proportion_Learner_Usage = \(-3.6489 + 0.8125 \times\) CGN_Freq + 0.3506 \* CELEX_Freq; AIC = 1308), compared to CELEX alone (AIC = 1363; \(p < 0.001\)), or
CGN alone (AIC = 1320; \( p < 0.001 \)). Although there is a difference in the fitted model parameters, it should be emphasized that CELEX frequencies did not greatly improve the model fit in the present analysis, as the difference in AIC between the full model and CGN-alone model is not large (1308 versus 1320).

This analysis revealed that for learner usage during a relatively early period of adult language learning, a combination of predictors is a more effective model of usage than any single predictor. In addition, the CGN word frequency and textbook status were stronger predictors than CELEX word frequencies. Without textbook status in the model, both CGN and CELEX were positive predictors, although again the relationship with CELEX was proportionately smaller.

**Discussion**

The analyses presented here revealed that both spoken (CGN) and written (CELEX) corpus frequency estimates predicted whether a word would be used as a vocabulary item in textbooks. However, spoken word frequencies predict best what learners are most likely to read in the classroom and including written frequency estimates did not result in a better prediction than spoken frequencies alone.

By contrast, in the analysis of actual (untutored) learner usage the model that predicted best whether speakers used a word in a film re-telling task included both spoken and written corpora and, surprisingly, even the textbook status of words. Given that our first analysis showed that all three sources are highly correlated, this result suggests that what makes combined estimates helpful is the variability in estimated frequency which is not shared among the sources. As a practical consequence, this result shows that no single word frequency count may be best for predicting learner usage or – assuming that usage and comprehension are highly correlated (Laufer et al., 2004) – learner vocabulary knowledge. Instead, a broad sample of many different language use domains should be used, if it is available. In cases where fewer sources are available, the use of textbook status and written frequency estimates alone appear problematic, but their combination was almost as good as spoken frequency
estimates. An interesting extension of the present analysis would be to use frequency of occurrence in a textbook as a predictor variable, if this information could be obtained in a convenient form.

Although the models were fitted to predict occurrence in learner speech about just one specific film topic, it is important to note that they also predict the use of words in other contexts. The words *tijd* “time”, *uur* “hour, o’clock”, *geval* “case”, *leven* “life”, *aantal* “number”, *jaren* “years”, *wereld* “world”, *manier* “manner”, *vraag* “question”, for example, are highly predicted although they are not used in the film re-tellings. This is not a flaw of the model. Simply speaking, the models find properties of words (e.g. to have a certain CGN frequency, to occur in textbooks) that make them likely to be used more often during the film re-tellings. Other words that have not been used during the film re-tellings may nonetheless have similar properties, so in a sense they are predicted to be used when talking about an appropriate topic. Of course, even in L2 learners the choice of words used is more strongly determined by what they are talking about than how well they know certain words. It is for the same reason that a broad sample of many different language use domains predicts learner usage best. A broad sample reflects better than any single source what people talk about in different contexts and therefore what is common in a language.

It is quite possible that the first language of an L2 learner could influence the likelihood that low-frequency L2 words are produced. If the L1 and L2 of a learner share cognates which happen to be low frequency, for example, then these words would be produced more often by that learner because of they would be already known (as cognates). A learner of the same L2 but with a different L1 that does not share the cognates would be less likely to produce these words.

Why are words that are more common in a language learned earlier than less common words? Perhaps the simplest explanation might be that commonly encountered words are simply needed more, other factors being equal, in order for effective communication to occur. Experimental research on memory has demonstrated that the so-called *need probability* of memory items (Anderson and Schooler, 1991) can be used to account for a variety of effects in research on memory including the shape of practice and forgetting functions, also including those of L2 lexical acquisition (Pavlik and Anderson, 2005). The need probability of an item is simply the probability that some item of information to be learned will be required again in the future. Lexical frequency can be seen as an estimate of the probability that a word will be required again sometime in the future. A productive line of research may be to relate how different aspects of lexical frequency, recency of encounter, and prior exposure over a learner’s lifespan may predict learner’s acquisition and retention of L2 words.

Several caveats to the approach we have taken are in order. In the present study, we considered whether two relatively large corpora would be useful in predicting learner usage. This might suggest that using even a larger number of data sources would improve predictions even more. However, the cost of obtaining each source has to be weighed against the potential decreasing utility of extending the number of data sources. It is sometimes time-consuming and expensive to obtain additional corpus data (especially with regard to speech data), and if there are already several data sources available, new sources may not add much more predictive ability. Second, if the data that are used are unrepresentative of the target language, it may not help with prediction. With these caveats in mind, although it is unlikely that word frequency alone can serve as a completely adequate predictor, the results presented here suggest that the probability that a word will be encountered in a learner’s environment can be most effectively modeled using multiple data sources.
## Appendix A. 200 Dutch lemmas with highest prediction of textbook appearance

<table>
<thead>
<tr>
<th>Lemma</th>
<th>FreqCGN</th>
<th>FreqCELEX</th>
<th>Ntext</th>
<th>Predicted</th>
<th>Lemma</th>
<th>FreqCGN</th>
<th>FreqCELEX</th>
<th>Ntext</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>jaar</td>
<td>1700</td>
<td>1143</td>
<td>3</td>
<td>0.992</td>
<td>politie</td>
<td>126</td>
<td>95</td>
<td>0</td>
<td>0.865</td>
</tr>
<tr>
<td>mens</td>
<td>1604</td>
<td>1370</td>
<td>3</td>
<td>0.991</td>
<td>brief</td>
<td>125</td>
<td>200</td>
<td>2</td>
<td>0.864</td>
</tr>
<tr>
<td>keer</td>
<td>1548</td>
<td>454</td>
<td>1</td>
<td>0.991</td>
<td>kaart</td>
<td>125</td>
<td>88</td>
<td>3</td>
<td>0.864</td>
</tr>
<tr>
<td>soort</td>
<td>1050</td>
<td>75</td>
<td>3</td>
<td>0.986</td>
<td>tent</td>
<td>124</td>
<td>27</td>
<td>1</td>
<td>0.863</td>
</tr>
<tr>
<td>zijn</td>
<td>1003</td>
<td>648</td>
<td>3</td>
<td>0.985</td>
<td>student</td>
<td>123</td>
<td>49</td>
<td>2</td>
<td>0.862</td>
</tr>
<tr>
<td>tijd</td>
<td>903</td>
<td>1084</td>
<td>3</td>
<td>0.983</td>
<td>grond</td>
<td>122</td>
<td>353</td>
<td>2</td>
<td>0.861</td>
</tr>
<tr>
<td>ding</td>
<td>874</td>
<td>371</td>
<td>3</td>
<td>0.983</td>
<td>maandag</td>
<td>122</td>
<td>16</td>
<td>3</td>
<td>0.861</td>
</tr>
<tr>
<td>week</td>
<td>834</td>
<td>283</td>
<td>3</td>
<td>0.982</td>
<td>mama</td>
<td>122</td>
<td>43</td>
<td>1</td>
<td>0.861</td>
</tr>
<tr>
<td>uur</td>
<td>808</td>
<td>425</td>
<td>3</td>
<td>0.981</td>
<td>straat</td>
<td>122</td>
<td>195</td>
<td>3</td>
<td>0.861</td>
</tr>
<tr>
<td>man</td>
<td>729</td>
<td>1190</td>
<td>3</td>
<td>0.979</td>
<td>muziek</td>
<td>121</td>
<td>115</td>
<td>3</td>
<td>0.860</td>
</tr>
<tr>
<td>kind</td>
<td>708</td>
<td>961</td>
<td>3</td>
<td>0.978</td>
<td>computer</td>
<td>120</td>
<td>49</td>
<td>2</td>
<td>0.858</td>
</tr>
<tr>
<td>punt</td>
<td>576</td>
<td>19</td>
<td>2</td>
<td>0.973</td>
<td>familie</td>
<td>120</td>
<td>134</td>
<td>2</td>
<td>0.858</td>
</tr>
<tr>
<td>moment</td>
<td>541</td>
<td>298</td>
<td>3</td>
<td>0.971</td>
<td>telefoon</td>
<td>120</td>
<td>84</td>
<td>3</td>
<td>0.858</td>
</tr>
<tr>
<td>stuk</td>
<td>532</td>
<td>46</td>
<td>3</td>
<td>0.970</td>
<td>buurt</td>
<td>118</td>
<td>109</td>
<td>3</td>
<td>0.856</td>
</tr>
<tr>
<td>idee</td>
<td>501</td>
<td>211</td>
<td>3</td>
<td>0.968</td>
<td>gebied</td>
<td>118</td>
<td>242</td>
<td>2</td>
<td>0.856</td>
</tr>
<tr>
<td>boek</td>
<td>499</td>
<td>387</td>
<td>3</td>
<td>0.968</td>
<td>orde</td>
<td>118</td>
<td>159</td>
<td>0</td>
<td>0.856</td>
</tr>
<tr>
<td>huis</td>
<td>495</td>
<td>630</td>
<td>3</td>
<td>0.968</td>
<td>wedstrijd</td>
<td>118</td>
<td>23</td>
<td>1</td>
<td>0.856</td>
</tr>
<tr>
<td>paai</td>
<td>488</td>
<td>491</td>
<td>2</td>
<td>0.967</td>
<td>kennis</td>
<td>117</td>
<td>141</td>
<td>3</td>
<td>0.855</td>
</tr>
<tr>
<td>geval</td>
<td>457</td>
<td>539</td>
<td>2</td>
<td>0.965</td>
<td>Amsterdam</td>
<td>115</td>
<td>95</td>
<td>0</td>
<td>0.852</td>
</tr>
<tr>
<td>jong</td>
<td>440</td>
<td>41</td>
<td>3</td>
<td>0.963</td>
<td>Belgie</td>
<td>114</td>
<td>33</td>
<td>1</td>
<td>0.851</td>
</tr>
<tr>
<td>vraag</td>
<td>423</td>
<td>476</td>
<td>3</td>
<td>0.962</td>
<td>gemeente</td>
<td>114</td>
<td>74</td>
<td>3</td>
<td>0.851</td>
</tr>
<tr>
<td>probleem</td>
<td>420</td>
<td>340</td>
<td>3</td>
<td>0.961</td>
<td>beeld</td>
<td>113</td>
<td>197</td>
<td>1</td>
<td>0.850</td>
</tr>
<tr>
<td>vrouw</td>
<td>404</td>
<td>900</td>
<td>3</td>
<td>0.960</td>
<td>antwoord</td>
<td>112</td>
<td>209</td>
<td>3</td>
<td>0.849</td>
</tr>
<tr>
<td>werk</td>
<td>398</td>
<td>571</td>
<td>2</td>
<td>0.959</td>
<td>gezicht</td>
<td>112</td>
<td>504</td>
<td>3</td>
<td>0.849</td>
</tr>
<tr>
<td>God</td>
<td>389</td>
<td>298</td>
<td>0</td>
<td>0.958</td>
<td>arm</td>
<td>111</td>
<td>187</td>
<td>1</td>
<td>0.847</td>
</tr>
<tr>
<td>school</td>
<td>388</td>
<td>243</td>
<td>3</td>
<td>0.958</td>
<td>reden</td>
<td>109</td>
<td>226</td>
<td>3</td>
<td>0.845</td>
</tr>
<tr>
<td>aantal</td>
<td>379</td>
<td>378</td>
<td>3</td>
<td>0.957</td>
<td>fiets</td>
<td>108</td>
<td>48</td>
<td>3</td>
<td>0.843</td>
</tr>
<tr>
<td>leerling</td>
<td>373</td>
<td>58</td>
<td>0</td>
<td>0.956</td>
<td>kilometer</td>
<td>108</td>
<td>65</td>
<td>2</td>
<td>0.843</td>
</tr>
<tr>
<td>Nederland</td>
<td>347</td>
<td>220</td>
<td>2</td>
<td>0.952</td>
<td>tekst</td>
<td>108</td>
<td>85</td>
<td>3</td>
<td>0.843</td>
</tr>
<tr>
<td>moeder</td>
<td>343</td>
<td>596</td>
<td>3</td>
<td>0.952</td>
<td>bedoeling</td>
<td>107</td>
<td>84</td>
<td>0</td>
<td>0.842</td>
</tr>
<tr>
<td>hand</td>
<td>342</td>
<td>1028</td>
<td>3</td>
<td>0.952</td>
<td>helft</td>
<td>107</td>
<td>82</td>
<td>3</td>
<td>0.842</td>
</tr>
<tr>
<td>kant</td>
<td>339</td>
<td>291</td>
<td>2</td>
<td>0.951</td>
<td>voet</td>
<td>107</td>
<td>225</td>
<td>3</td>
<td>0.842</td>
</tr>
<tr>
<td>Word</td>
<td>Frequency</td>
<td>Position</td>
<td>Length</td>
<td>Confidence</td>
<td>Length</td>
<td>Position</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>----------</td>
<td>--------</td>
<td>------------</td>
<td>--------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>manier</td>
<td>336</td>
<td>3</td>
<td>3</td>
<td>0.951</td>
<td>vak</td>
<td>105</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>minuut</td>
<td>327</td>
<td>3</td>
<td>3</td>
<td>0.949</td>
<td>gang</td>
<td>104</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>plaats</td>
<td>313</td>
<td>3</td>
<td>3</td>
<td>0.947</td>
<td>programma</td>
<td>104</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>geld</td>
<td>311</td>
<td>3</td>
<td>3</td>
<td>0.947</td>
<td>vorm</td>
<td>104</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>maand</td>
<td>305</td>
<td>3</td>
<td>3</td>
<td>0.945</td>
<td>ruimte</td>
<td>102</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zin</td>
<td>305</td>
<td>2</td>
<td>3</td>
<td>0.945</td>
<td>voorbeeld</td>
<td>102</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>auto</td>
<td>300</td>
<td>3</td>
<td>3</td>
<td>0.944</td>
<td>kerk</td>
<td>101</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>heer</td>
<td>297</td>
<td>1</td>
<td>3</td>
<td>0.944</td>
<td>donderdag</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vader</td>
<td>282</td>
<td>3</td>
<td>3</td>
<td>0.941</td>
<td>stem</td>
<td>99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>leven</td>
<td>279</td>
<td>3</td>
<td>3</td>
<td>0.940</td>
<td>einde</td>
<td>98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>minister</td>
<td>279</td>
<td>0</td>
<td>3</td>
<td>0.940</td>
<td>dinsdag</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>avond</td>
<td>278</td>
<td>3</td>
<td>3</td>
<td>0.940</td>
<td>onderwijs</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bal</td>
<td>278</td>
<td>2</td>
<td>3</td>
<td>0.940</td>
<td>moeite</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eind</td>
<td>278</td>
<td>1</td>
<td>3</td>
<td>0.940</td>
<td>mogelijkheid</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>meneer</td>
<td>266</td>
<td>2</td>
<td>3</td>
<td>0.937</td>
<td>woensdag</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>groep</td>
<td>263</td>
<td>3</td>
<td>3</td>
<td>0.936</td>
<td>Frankrijk</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zaak</td>
<td>252</td>
<td>2</td>
<td>3</td>
<td>0.933</td>
<td>Brussel</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>naam</td>
<td>239</td>
<td>3</td>
<td>3</td>
<td>0.929</td>
<td>situatie</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stad</td>
<td>236</td>
<td>3</td>
<td>3</td>
<td>0.928</td>
<td>dorp</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kamer</td>
<td>223</td>
<td>3</td>
<td>3</td>
<td>0.924</td>
<td>baan</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>oog</td>
<td>220</td>
<td>3</td>
<td>3</td>
<td>0.923</td>
<td>ogenblik</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>verhaal</td>
<td>217</td>
<td>2</td>
<td>3</td>
<td>0.922</td>
<td>tuin</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hoop</td>
<td>216</td>
<td>1</td>
<td>3</td>
<td>0.921</td>
<td>hond</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>water</td>
<td>215</td>
<td>3</td>
<td>3</td>
<td>0.921</td>
<td>papa</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>feit</td>
<td>211</td>
<td>1</td>
<td>3</td>
<td>0.920</td>
<td>bank</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gulden</td>
<td>210</td>
<td>1</td>
<td>3</td>
<td>0.919</td>
<td>feest</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>begin</td>
<td>202</td>
<td>2</td>
<td>3</td>
<td>0.916</td>
<td>persoon</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ouder</td>
<td>201</td>
<td>3</td>
<td>3</td>
<td>0.915</td>
<td>licht</td>
<td>87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>procent</td>
<td>199</td>
<td>2</td>
<td>3</td>
<td>0.915</td>
<td>rol</td>
<td>87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>land</td>
<td>196</td>
<td>3</td>
<td>3</td>
<td>0.913</td>
<td>aandacht</td>
<td>86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gesprek</td>
<td>190</td>
<td>3</td>
<td>3</td>
<td>0.910</td>
<td>kleed</td>
<td>86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hoofd</td>
<td>186</td>
<td>3</td>
<td>3</td>
<td>0.908</td>
<td>niveau</td>
<td>86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mevrouw</td>
<td>185</td>
<td>3</td>
<td>3</td>
<td>0.908</td>
<td>periode</td>
<td>86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lemma</td>
<td>FreqCGN</td>
<td>FreqCELEX</td>
<td>Ntext</td>
<td>Predicted</td>
<td>Lemma</td>
<td>FreqCGN</td>
<td>FreqCELEX</td>
<td>Ntext</td>
<td>Predicted</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>-----------</td>
<td>-------</td>
<td>-----------</td>
<td>--------</td>
<td>---------</td>
<td>-----------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>rest</td>
<td>185</td>
<td>115</td>
<td>2</td>
<td>0.908</td>
<td>reis</td>
<td>86</td>
<td>98</td>
<td>3</td>
<td>0.807</td>
</tr>
<tr>
<td>klas</td>
<td>180</td>
<td>48</td>
<td>2</td>
<td>0.905</td>
<td>plek</td>
<td>85</td>
<td>90</td>
<td>1</td>
<td>0.805</td>
</tr>
<tr>
<td>meisje</td>
<td>180</td>
<td>357</td>
<td>3</td>
<td>0.905</td>
<td>zoon</td>
<td>85</td>
<td>189</td>
<td>3</td>
<td>0.805</td>
</tr>
<tr>
<td>vriend</td>
<td>178</td>
<td>284</td>
<td>3</td>
<td>0.904</td>
<td>a</td>
<td>84</td>
<td>47</td>
<td>0</td>
<td>0.803</td>
</tr>
<tr>
<td>film</td>
<td>174</td>
<td>106</td>
<td>3</td>
<td>0.902</td>
<td>onderwerp</td>
<td>84</td>
<td>93</td>
<td>2</td>
<td>0.803</td>
</tr>
<tr>
<td>weekend</td>
<td>173</td>
<td>17</td>
<td>3</td>
<td>0.901</td>
<td>vriendin</td>
<td>84</td>
<td>72</td>
<td>3</td>
<td>0.803</td>
</tr>
<tr>
<td>wereld</td>
<td>167</td>
<td>454</td>
<td>1</td>
<td>0.898</td>
<td>informatie</td>
<td>83</td>
<td>109</td>
<td>3</td>
<td>0.800</td>
</tr>
<tr>
<td>bus</td>
<td>166</td>
<td>40</td>
<td>3</td>
<td>0.897</td>
<td>wet</td>
<td>83</td>
<td>187</td>
<td>0</td>
<td>0.800</td>
</tr>
<tr>
<td>les</td>
<td>166</td>
<td>32</td>
<td>2</td>
<td>0.897</td>
<td>lijn</td>
<td>82</td>
<td>104</td>
<td>3</td>
<td>0.798</td>
</tr>
<tr>
<td>deel</td>
<td>164</td>
<td>388</td>
<td>2</td>
<td>0.896</td>
<td>muur</td>
<td>82</td>
<td>147</td>
<td>3</td>
<td>0.798</td>
</tr>
<tr>
<td>deur</td>
<td>162</td>
<td>376</td>
<td>3</td>
<td>0.895</td>
<td>nieuws</td>
<td>82</td>
<td>29</td>
<td>3</td>
<td>0.798</td>
</tr>
<tr>
<td>onderzoek</td>
<td>160</td>
<td>204</td>
<td>1</td>
<td>0.893</td>
<td>principe</td>
<td>82</td>
<td>77</td>
<td>1</td>
<td>0.798</td>
</tr>
<tr>
<td>foto</td>
<td>155</td>
<td>107</td>
<td>3</td>
<td>0.890</td>
<td>winkel</td>
<td>82</td>
<td>60</td>
<td>3</td>
<td>0.798</td>
</tr>
<tr>
<td>voorzitter</td>
<td>155</td>
<td>48</td>
<td>0</td>
<td>0.890</td>
<td>blik</td>
<td>81</td>
<td>188</td>
<td>1</td>
<td>0.796</td>
</tr>
<tr>
<td>bedrijf</td>
<td>153</td>
<td>119</td>
<td>3</td>
<td>0.888</td>
<td>eeuw</td>
<td>81</td>
<td>229</td>
<td>2</td>
<td>0.796</td>
</tr>
<tr>
<td>jongen</td>
<td>151</td>
<td>360</td>
<td>2</td>
<td>0.887</td>
<td>Frans</td>
<td>81</td>
<td>43</td>
<td>1</td>
<td>0.796</td>
</tr>
<tr>
<td>nacht</td>
<td>151</td>
<td>266</td>
<td>3</td>
<td>0.887</td>
<td>ziekenhuis</td>
<td>81</td>
<td>94</td>
<td>3</td>
<td>0.796</td>
</tr>
<tr>
<td>vrijdag</td>
<td>151</td>
<td>17</td>
<td>3</td>
<td>0.887</td>
<td>zon</td>
<td>81</td>
<td>46</td>
<td>3</td>
<td>0.796</td>
</tr>
<tr>
<td>zaterdag</td>
<td>146</td>
<td>21</td>
<td>3</td>
<td>0.883</td>
<td>dochter</td>
<td>80</td>
<td>120</td>
<td>3</td>
<td>0.794</td>
</tr>
<tr>
<td>kans</td>
<td>144</td>
<td>202</td>
<td>0</td>
<td>0.882</td>
<td>indruk</td>
<td>80</td>
<td>155</td>
<td>2</td>
<td>0.794</td>
</tr>
<tr>
<td>ure</td>
<td>140</td>
<td>39</td>
<td>0</td>
<td>0.878</td>
<td>rug</td>
<td>80</td>
<td>180</td>
<td>3</td>
<td>0.794</td>
</tr>
<tr>
<td>bed</td>
<td>139</td>
<td>300</td>
<td>3</td>
<td>0.877</td>
<td>staat</td>
<td>80</td>
<td>290</td>
<td>0</td>
<td>0.794</td>
</tr>
<tr>
<td>trein</td>
<td>138</td>
<td>81</td>
<td>3</td>
<td>0.876</td>
<td>Engels</td>
<td>79</td>
<td>35</td>
<td>1</td>
<td>0.791</td>
</tr>
<tr>
<td>verschil</td>
<td>138</td>
<td>175</td>
<td>3</td>
<td>0.876</td>
<td>overheid</td>
<td>79</td>
<td>127</td>
<td>1</td>
<td>0.791</td>
</tr>
<tr>
<td>partij</td>
<td>136</td>
<td>170</td>
<td>0</td>
<td>0.875</td>
<td>rekening</td>
<td>79</td>
<td>114</td>
<td>2</td>
<td>0.791</td>
</tr>
<tr>
<td>weer</td>
<td>136</td>
<td>16</td>
<td>3</td>
<td>0.875</td>
<td>steen</td>
<td>79</td>
<td>12</td>
<td>1</td>
<td>0.791</td>
</tr>
<tr>
<td>tafel</td>
<td>135</td>
<td>247</td>
<td>3</td>
<td>0.874</td>
<td>afspraak</td>
<td>78</td>
<td>53</td>
<td>2</td>
<td>0.789</td>
</tr>
<tr>
<td>zondag</td>
<td>131</td>
<td>41</td>
<td>3</td>
<td>0.870</td>
<td>kop</td>
<td>78</td>
<td>135</td>
<td>2</td>
<td>0.789</td>
</tr>
<tr>
<td>regering</td>
<td>130</td>
<td>115</td>
<td>0</td>
<td>0.869</td>
<td>papier</td>
<td>78</td>
<td>113</td>
<td>3</td>
<td>0.789</td>
</tr>
<tr>
<td>taal</td>
<td>130</td>
<td>156</td>
<td>2</td>
<td>0.869</td>
<td>stof</td>
<td>78</td>
<td>6</td>
<td>0</td>
<td>0.789</td>
</tr>
<tr>
<td>richting</td>
<td>129</td>
<td>199</td>
<td>2</td>
<td>0.868</td>
<td>been</td>
<td>77</td>
<td>178</td>
<td>3</td>
<td>0.787</td>
</tr>
<tr>
<td>broer</td>
<td>128</td>
<td>128</td>
<td>3</td>
<td>0.867</td>
<td>eten</td>
<td>77</td>
<td>103</td>
<td>2</td>
<td>0.787</td>
</tr>
<tr>
<td>gevoel</td>
<td>128</td>
<td>251</td>
<td>3</td>
<td>0.867</td>
<td>raam</td>
<td>77</td>
<td>174</td>
<td>3</td>
<td>0.787</td>
</tr>
<tr>
<td>krant</td>
<td>128</td>
<td>117</td>
<td>3</td>
<td>0.867</td>
<td>lichaam</td>
<td>75</td>
<td>292</td>
<td>2</td>
<td>0.782</td>
</tr>
<tr>
<td>nummer</td>
<td>128</td>
<td>72</td>
<td>3</td>
<td>0.867</td>
<td>Nederland</td>
<td>75</td>
<td>35</td>
<td>2</td>
<td>0.782</td>
</tr>
</tbody>
</table>
### Appendix B. 200 Dutch word forms with highest prediction of appearance in learner production

<table>
<thead>
<tr>
<th>Form</th>
<th>CGN</th>
<th>CELEX</th>
<th>NTxt</th>
<th>Film</th>
<th>Predicted</th>
<th>Form</th>
<th>CGN</th>
<th>CELEX</th>
<th>NTxt</th>
<th>Film</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>mensen</td>
<td>1469</td>
<td>934</td>
<td>3</td>
<td>1</td>
<td>0.7131</td>
<td>reden</td>
<td>86</td>
<td>164</td>
<td>3</td>
<td>0</td>
<td>0.2357</td>
</tr>
<tr>
<td>tijd</td>
<td>807</td>
<td>958</td>
<td>3</td>
<td>0</td>
<td>0.6802</td>
<td>hart</td>
<td>61</td>
<td>183</td>
<td>2</td>
<td>0</td>
<td>0.2323</td>
</tr>
<tr>
<td>jaar</td>
<td>1438</td>
<td>734</td>
<td>3</td>
<td>1</td>
<td>0.6746</td>
<td>vragen</td>
<td>130</td>
<td>134</td>
<td>3</td>
<td>1</td>
<td>0.2309</td>
</tr>
<tr>
<td>man</td>
<td>556</td>
<td>876</td>
<td>3</td>
<td>1</td>
<td>0.6414</td>
<td>verband</td>
<td>64</td>
<td>177</td>
<td>1</td>
<td>0</td>
<td>0.2304</td>
</tr>
<tr>
<td>zijn</td>
<td>1003</td>
<td>648</td>
<td>3</td>
<td>1</td>
<td>0.6303</td>
<td>straat</td>
<td>102</td>
<td>147</td>
<td>3</td>
<td>1</td>
<td>0.2304</td>
</tr>
<tr>
<td>keer</td>
<td>1493</td>
<td>426</td>
<td>1</td>
<td>1</td>
<td>0.5852</td>
<td>blik</td>
<td>76</td>
<td>164</td>
<td>1</td>
<td>0</td>
<td>0.2294</td>
</tr>
<tr>
<td>dag</td>
<td>600</td>
<td>607</td>
<td>2</td>
<td>1</td>
<td>0.5834</td>
<td>kennis</td>
<td>108</td>
<td>141</td>
<td>3</td>
<td>0</td>
<td>0.2280</td>
</tr>
<tr>
<td>huis</td>
<td>431</td>
<td>541</td>
<td>3</td>
<td>1</td>
<td>0.5391</td>
<td>moeite</td>
<td>95</td>
<td>148</td>
<td>1</td>
<td>0</td>
<td>0.2276</td>
</tr>
<tr>
<td>kinderen</td>
<td>491</td>
<td>484</td>
<td>3</td>
<td>1</td>
<td>0.5284</td>
<td>boeken</td>
<td>151</td>
<td>121</td>
<td>3</td>
<td>0</td>
<td>0.2255</td>
</tr>
<tr>
<td>paar</td>
<td>486</td>
<td>483</td>
<td>2</td>
<td>1</td>
<td>0.5273</td>
<td>jongens</td>
<td>133</td>
<td>127</td>
<td>2</td>
<td>1</td>
<td>0.2252</td>
</tr>
<tr>
<td>moeder</td>
<td>325</td>
<td>555</td>
<td>3</td>
<td>1</td>
<td>0.5233</td>
<td>oog</td>
<td>60</td>
<td>171</td>
<td>3</td>
<td>0</td>
<td>0.2228</td>
</tr>
<tr>
<td>hand</td>
<td>215</td>
<td>645</td>
<td>3</td>
<td>1</td>
<td>0.5207</td>
<td>ruimte</td>
<td>100</td>
<td>139</td>
<td>2</td>
<td>0</td>
<td>0.2223</td>
</tr>
<tr>
<td>vrouw</td>
<td>258</td>
<td>597</td>
<td>3</td>
<td>1</td>
<td>0.5199</td>
<td>plan</td>
<td>98</td>
<td>137</td>
<td>3</td>
<td>0</td>
<td>0.2195</td>
</tr>
<tr>
<td>plaats</td>
<td>264</td>
<td>590</td>
<td>3</td>
<td>1</td>
<td>0.5194</td>
<td>gesprek</td>
<td>147</td>
<td>116</td>
<td>3</td>
<td>0</td>
<td>0.2189</td>
</tr>
<tr>
<td>uur</td>
<td>744</td>
<td>380</td>
<td>3</td>
<td>0</td>
<td>0.5147</td>
<td>taal</td>
<td>109</td>
<td>130</td>
<td>2</td>
<td>0</td>
<td>0.2182</td>
</tr>
<tr>
<td>werk</td>
<td>380</td>
<td>496</td>
<td>2</td>
<td>1</td>
<td>0.5143</td>
<td>zoen</td>
<td>72</td>
<td>151</td>
<td>3</td>
<td>0</td>
<td>0.2164</td>
</tr>
<tr>
<td>vader</td>
<td>271</td>
<td>547</td>
<td>3</td>
<td>1</td>
<td>0.5075</td>
<td>beeld</td>
<td>77</td>
<td>146</td>
<td>1</td>
<td>0</td>
<td>0.2155</td>
</tr>
<tr>
<td>ogen</td>
<td>160</td>
<td>649</td>
<td>3</td>
<td>1</td>
<td>0.5004</td>
<td>pijn</td>
<td>71</td>
<td>149</td>
<td>2</td>
<td>1</td>
<td>0.2140</td>
</tr>
<tr>
<td>geval</td>
<td>425</td>
<td>411</td>
<td>2</td>
<td>0</td>
<td>0.4882</td>
<td>indruk</td>
<td>78</td>
<td>143</td>
<td>2</td>
<td>0</td>
<td>0.2136</td>
</tr>
<tr>
<td>leven</td>
<td>277</td>
<td>443</td>
<td>3</td>
<td>0</td>
<td>0.4708</td>
<td>meter</td>
<td>153</td>
<td>108</td>
<td>2</td>
<td>1</td>
<td>0.2122</td>
</tr>
<tr>
<td>dingen</td>
<td>724</td>
<td>298</td>
<td>3</td>
<td>1</td>
<td>0.4686</td>
<td>familie</td>
<td>112</td>
<td>122</td>
<td>2</td>
<td>1</td>
<td>0.2118</td>
</tr>
<tr>
<td>hoofd</td>
<td>179</td>
<td>515</td>
<td>3</td>
<td>1</td>
<td>0.4666</td>
<td>vriend</td>
<td>74</td>
<td>143</td>
<td>3</td>
<td>1</td>
<td>0.2111</td>
</tr>
<tr>
<td>aantal</td>
<td>371</td>
<td>364</td>
<td>3</td>
<td>0</td>
<td>0.4564</td>
<td>liefde</td>
<td>49</td>
<td>168</td>
<td>1</td>
<td>0</td>
<td>0.2108</td>
</tr>
<tr>
<td>jaren</td>
<td>262</td>
<td>406</td>
<td>3</td>
<td>0</td>
<td>0.4511</td>
<td>rest</td>
<td>181</td>
<td>98</td>
<td>2</td>
<td>0</td>
<td>0.2086</td>
</tr>
<tr>
<td>kind</td>
<td>196</td>
<td>454</td>
<td>3</td>
<td>1</td>
<td>0.4503</td>
<td>muziek</td>
<td>121</td>
<td>115</td>
<td>3</td>
<td>0</td>
<td>0.2084</td>
</tr>
<tr>
<td>wereld</td>
<td>164</td>
<td>444</td>
<td>1</td>
<td>0</td>
<td>0.4336</td>
<td>begin</td>
<td>202</td>
<td>91</td>
<td>2</td>
<td>1</td>
<td>0.2050</td>
</tr>
<tr>
<td>manier</td>
<td>316</td>
<td>333</td>
<td>3</td>
<td>0</td>
<td>0.4290</td>
<td>geschiedenis</td>
<td>70</td>
<td>135</td>
<td>2</td>
<td>0</td>
<td>0.2016</td>
</tr>
<tr>
<td>vraag</td>
<td>293</td>
<td>342</td>
<td>3</td>
<td>0</td>
<td>0.4284</td>
<td>toekomst</td>
<td>74</td>
<td>131</td>
<td>1</td>
<td>0</td>
<td>0.2007</td>
</tr>
<tr>
<td>moment</td>
<td>516</td>
<td>270</td>
<td>3</td>
<td>1</td>
<td>0.4266</td>
<td>zee</td>
<td>64</td>
<td>138</td>
<td>3</td>
<td>1</td>
<td>0.2001</td>
</tr>
<tr>
<td>mens</td>
<td>135</td>
<td>436</td>
<td>3</td>
<td>1</td>
<td>0.4166</td>
<td>hoop</td>
<td>216</td>
<td>84</td>
<td>1</td>
<td>0</td>
<td>0.1989</td>
</tr>
<tr>
<td>water</td>
<td>213</td>
<td>353</td>
<td>3</td>
<td>1</td>
<td>0.4114</td>
<td>vrienden</td>
<td>80</td>
<td>125</td>
<td>3</td>
<td>0</td>
<td>0.1989</td>
</tr>
<tr>
<td>zin</td>
<td>284</td>
<td>312</td>
<td>2</td>
<td>0</td>
<td>0.4099</td>
<td>verschil</td>
<td>116</td>
<td>107</td>
<td>3</td>
<td>0</td>
<td>0.1980</td>
</tr>
<tr>
<td>weg</td>
<td>158</td>
<td>386</td>
<td>3</td>
<td>1</td>
<td>0.4062</td>
<td>haar</td>
<td>934</td>
<td>46</td>
<td>3</td>
<td>1</td>
<td>0.1979</td>
</tr>
<tr>
<td>Form</td>
<td>CGN</td>
<td>CELEX</td>
<td>NTxt</td>
<td>Film</td>
<td>Predicted</td>
<td>Form</td>
<td>CGN</td>
<td>CELEX</td>
<td>NTxt</td>
<td>Film</td>
<td>Predicted</td>
</tr>
<tr>
<td>--------</td>
<td>-----</td>
<td>-------</td>
<td>------</td>
<td>------</td>
<td>-----------</td>
<td>--------</td>
<td>-----</td>
<td>-------</td>
<td>------</td>
<td>------</td>
<td>-----------</td>
</tr>
<tr>
<td>gezicht</td>
<td>99</td>
<td>448</td>
<td>3</td>
<td>0</td>
<td>0.3998</td>
<td>systeem</td>
<td>60</td>
<td>138</td>
<td>1</td>
<td>0</td>
<td>0.1971</td>
</tr>
<tr>
<td>geld</td>
<td>311</td>
<td>276</td>
<td>3</td>
<td>1</td>
<td>0.3949</td>
<td>buurt</td>
<td>118</td>
<td>105</td>
<td>3</td>
<td>0</td>
<td>0.1967</td>
</tr>
<tr>
<td>kamer</td>
<td>180</td>
<td>318</td>
<td>3</td>
<td>1</td>
<td>0.3816</td>
<td>wijn</td>
<td>53</td>
<td>141</td>
<td>3</td>
<td>0</td>
<td>0.1940</td>
</tr>
<tr>
<td>handen</td>
<td>117</td>
<td>377</td>
<td>3</td>
<td>1</td>
<td>0.3814</td>
<td>periode</td>
<td>81</td>
<td>119</td>
<td>2</td>
<td>0</td>
<td>0.1938</td>
</tr>
<tr>
<td>land</td>
<td>196</td>
<td>298</td>
<td>3</td>
<td>0</td>
<td>0.3763</td>
<td>landen</td>
<td>85</td>
<td>114</td>
<td>3</td>
<td>0</td>
<td>0.1912</td>
</tr>
<tr>
<td>deur</td>
<td>139</td>
<td>325</td>
<td>3</td>
<td>1</td>
<td>0.3678</td>
<td>dokter</td>
<td>60</td>
<td>130</td>
<td>3</td>
<td>0</td>
<td>0.1904</td>
</tr>
<tr>
<td>boek</td>
<td>262</td>
<td>250</td>
<td>3</td>
<td>0</td>
<td>0.3662</td>
<td>aarde</td>
<td>37</td>
<td>153</td>
<td>1</td>
<td>0</td>
<td>0.1873</td>
</tr>
<tr>
<td>woorden</td>
<td>152</td>
<td>310</td>
<td>3</td>
<td>0</td>
<td>0.3658</td>
<td>brief</td>
<td>77</td>
<td>114</td>
<td>2</td>
<td>1</td>
<td>0.1868</td>
</tr>
<tr>
<td>naam</td>
<td>170</td>
<td>294</td>
<td>3</td>
<td>0</td>
<td>0.3644</td>
<td>kop</td>
<td>71</td>
<td>117</td>
<td>2</td>
<td>1</td>
<td>0.1861</td>
</tr>
<tr>
<td>woord</td>
<td>188</td>
<td>282</td>
<td>3</td>
<td>0</td>
<td>0.3642</td>
<td>overheid</td>
<td>71</td>
<td>117</td>
<td>1</td>
<td>0</td>
<td>0.1861</td>
</tr>
<tr>
<td>kant</td>
<td>295</td>
<td>235</td>
<td>2</td>
<td>1</td>
<td>0.3638</td>
<td>voeten</td>
<td>54</td>
<td>129</td>
<td>3</td>
<td>0</td>
<td>0.1849</td>
</tr>
<tr>
<td>deel</td>
<td>143</td>
<td>312</td>
<td>2</td>
<td>0</td>
<td>0.3628</td>
<td>meisjes</td>
<td>61</td>
<td>120</td>
<td>3</td>
<td>1</td>
<td>0.1823</td>
</tr>
<tr>
<td>mannen</td>
<td>145</td>
<td>296</td>
<td>3</td>
<td>0</td>
<td>0.3550</td>
<td>informatie</td>
<td>83</td>
<td>106</td>
<td>3</td>
<td>0</td>
<td>0.1821</td>
</tr>
<tr>
<td>grond</td>
<td>117</td>
<td>321</td>
<td>2</td>
<td>1</td>
<td>0.3543</td>
<td>beweging</td>
<td>38</td>
<td>144</td>
<td>1</td>
<td>0</td>
<td>0.1818</td>
</tr>
<tr>
<td>Nederland</td>
<td>346</td>
<td>202</td>
<td>2</td>
<td>0</td>
<td>0.3492</td>
<td>dienst</td>
<td>54</td>
<td>125</td>
<td>2</td>
<td>0</td>
<td>0.1815</td>
</tr>
<tr>
<td>school</td>
<td>345</td>
<td>198</td>
<td>3</td>
<td>0</td>
<td>0.3458</td>
<td>koffie</td>
<td>72</td>
<td>111</td>
<td>3</td>
<td>1</td>
<td>0.1810</td>
</tr>
<tr>
<td>stad</td>
<td>188</td>
<td>249</td>
<td>3</td>
<td>0</td>
<td>0.3435</td>
<td>werkelijkheid</td>
<td>32</td>
<td>152</td>
<td>1</td>
<td>0</td>
<td>0.1804</td>
</tr>
<tr>
<td>week</td>
<td>609</td>
<td>155</td>
<td>3</td>
<td>0</td>
<td>0.3430</td>
<td>rekening</td>
<td>75</td>
<td>107</td>
<td>2</td>
<td>0</td>
<td>0.1788</td>
</tr>
<tr>
<td>bed</td>
<td>133</td>
<td>284</td>
<td>3</td>
<td>0</td>
<td>0.3425</td>
<td>staat</td>
<td>70</td>
<td>273</td>
<td>0</td>
<td>1</td>
<td>0.1783</td>
</tr>
<tr>
<td>vrouwen</td>
<td>128</td>
<td>288</td>
<td>3</td>
<td>0</td>
<td>0.3423</td>
<td>eten</td>
<td>77</td>
<td>103</td>
<td>2</td>
<td>1</td>
<td>0.1759</td>
</tr>
<tr>
<td>meneer</td>
<td>266</td>
<td>203</td>
<td>2</td>
<td>1</td>
<td>0.3329</td>
<td>beleid</td>
<td>51</td>
<td>121</td>
<td>1</td>
<td>0</td>
<td>0.1757</td>
</tr>
<tr>
<td>gevoel</td>
<td>128</td>
<td>251</td>
<td>3</td>
<td>0</td>
<td>0.3202</td>
<td>benen</td>
<td>42</td>
<td>130</td>
<td>3</td>
<td>0</td>
<td>0.1752</td>
</tr>
<tr>
<td>ouders</td>
<td>193</td>
<td>206</td>
<td>3</td>
<td>0</td>
<td>0.3149</td>
<td>film</td>
<td>142</td>
<td>80</td>
<td>3</td>
<td>1</td>
<td>0.1751</td>
</tr>
<tr>
<td>idee</td>
<td>468</td>
<td>144</td>
<td>3</td>
<td>1</td>
<td>0.3144</td>
<td>tuin</td>
<td>84</td>
<td>98</td>
<td>3</td>
<td>1</td>
<td>0.1744</td>
</tr>
<tr>
<td>groep</td>
<td>180</td>
<td>201</td>
<td>3</td>
<td>0</td>
<td>0.3067</td>
<td>glas</td>
<td>44</td>
<td>124</td>
<td>1</td>
<td>0</td>
<td>0.1722</td>
</tr>
<tr>
<td>meisje</td>
<td>119</td>
<td>237</td>
<td>3</td>
<td>1</td>
<td>0.3067</td>
<td>maand</td>
<td>156</td>
<td>74</td>
<td>3</td>
<td>0</td>
<td>0.1710</td>
</tr>
<tr>
<td>onderzoek</td>
<td>160</td>
<td>204</td>
<td>1</td>
<td>0</td>
<td>0.3018</td>
<td>avonds</td>
<td>144</td>
<td>76</td>
<td>3</td>
<td>0</td>
<td>0.1704</td>
</tr>
<tr>
<td>licht</td>
<td>80</td>
<td>268</td>
<td>1</td>
<td>0</td>
<td>0.3013</td>
<td>persoon</td>
<td>59</td>
<td>108</td>
<td>2</td>
<td>1</td>
<td>0.1699</td>
</tr>
<tr>
<td>auto</td>
<td>256</td>
<td>165</td>
<td>3</td>
<td>1</td>
<td>0.2981</td>
<td>afstand</td>
<td>48</td>
<td>117</td>
<td>3</td>
<td>0</td>
<td>0.1697</td>
</tr>
<tr>
<td>stem</td>
<td>77</td>
<td>263</td>
<td>1</td>
<td>0</td>
<td>0.2962</td>
<td>mogelijkheid</td>
<td>53</td>
<td>112</td>
<td>1</td>
<td>0</td>
<td>0.1693</td>
</tr>
<tr>
<td>probleem</td>
<td>273</td>
<td>158</td>
<td>3</td>
<td>1</td>
<td>0.2954</td>
<td>ervaring</td>
<td>54</td>
<td>111</td>
<td>2</td>
<td>0</td>
<td>0.1691</td>
</tr>
<tr>
<td>lichaam</td>
<td>69</td>
<td>264</td>
<td>2</td>
<td>0</td>
<td>0.2902</td>
<td>feite</td>
<td>82</td>
<td>94</td>
<td>1</td>
<td>0</td>
<td>0.1691</td>
</tr>
<tr>
<td>Likely L2 words</td>
<td>Count</td>
<td>Index</td>
<td>Position</td>
<td>TF-IDF</td>
<td>TF-IDF Threshold</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>-------</td>
<td>-------</td>
<td>----------</td>
<td>--------</td>
<td>-----------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wijze</td>
<td>65</td>
<td>268</td>
<td>1</td>
<td>0.2889</td>
<td>0.1690</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vorm</td>
<td>84</td>
<td>242</td>
<td>3</td>
<td>0.2889</td>
<td>0.1674</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zaken</td>
<td>127</td>
<td>201</td>
<td>2</td>
<td>0.2857</td>
<td>0.1672</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zaak</td>
<td>115</td>
<td>204</td>
<td>2</td>
<td>0.2821</td>
<td>0.1666</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>avond</td>
<td>124</td>
<td>195</td>
<td>3</td>
<td>0.2798</td>
<td>0.1659</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>problemen</td>
<td>147</td>
<td>180</td>
<td>3</td>
<td>0.2781</td>
<td>0.1654</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mevrouw</td>
<td>185</td>
<td>164</td>
<td>3</td>
<td>1</td>
<td>0.2779</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>feit</td>
<td>108</td>
<td>203</td>
<td>1</td>
<td>0</td>
<td>0.2777</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>minuten</td>
<td>281</td>
<td>136</td>
<td>3</td>
<td>0</td>
<td>0.2750</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dagen</td>
<td>306</td>
<td>325</td>
<td>0</td>
<td>1</td>
<td>0.2739</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>God</td>
<td>376</td>
<td>298</td>
<td>0</td>
<td>0</td>
<td>0.2733</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tafel</td>
<td>115</td>
<td>189</td>
<td>3</td>
<td>1</td>
<td>0.2710</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>verhaal</td>
<td>165</td>
<td>161</td>
<td>2</td>
<td>1</td>
<td>0.2687</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>richting</td>
<td>119</td>
<td>183</td>
<td>2</td>
<td>0</td>
<td>0.2683</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mond</td>
<td>70</td>
<td>220</td>
<td>2</td>
<td>0</td>
<td>0.2645</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aandacht</td>
<td>86</td>
<td>199</td>
<td>2</td>
<td>0</td>
<td>0.2619</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dood</td>
<td>57</td>
<td>234</td>
<td>3</td>
<td>1</td>
<td>0.2618</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nacht</td>
<td>89</td>
<td>191</td>
<td>3</td>
<td>0</td>
<td>0.2581</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>maanden</td>
<td>149</td>
<td>155</td>
<td>3</td>
<td>1</td>
<td>0.2576</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gebied</td>
<td>96</td>
<td>183</td>
<td>2</td>
<td>0</td>
<td>0.2563</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>weken</td>
<td>225</td>
<td>128</td>
<td>3</td>
<td>0</td>
<td>0.2540</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>antwoord</td>
<td>93</td>
<td>177</td>
<td>3</td>
<td>0</td>
<td>0.2500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gebruik</td>
<td>64</td>
<td>205</td>
<td>1</td>
<td>0</td>
<td>0.2499</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rol</td>
<td>81</td>
<td>183</td>
<td>2</td>
<td>0</td>
<td>0.2471</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ogenblik</td>
<td>88</td>
<td>177</td>
<td>1</td>
<td>0</td>
<td>0.2471</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>situatie</td>
<td>76</td>
<td>185</td>
<td>3</td>
<td>0</td>
<td>0.2452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gang</td>
<td>97</td>
<td>167</td>
<td>3</td>
<td>0</td>
<td>0.2445</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>heer</td>
<td>258</td>
<td>111</td>
<td>1</td>
<td>0</td>
<td>0.2422</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eeuw</td>
<td>69</td>
<td>186</td>
<td>2</td>
<td>0</td>
<td>0.2408</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kerk</td>
<td>86</td>
<td>170</td>
<td>3</td>
<td>0</td>
<td>0.2405</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>oorlog</td>
<td>68</td>
<td>184</td>
<td>2</td>
<td>0</td>
<td>0.2386</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>voorbeeld</td>
<td>77</td>
<td>175</td>
<td>2</td>
<td>0</td>
<td>0.2385</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rug</td>
<td>78</td>
<td>173</td>
<td>3</td>
<td>0</td>
<td>0.2377</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lucht</td>
<td>69</td>
<td>181</td>
<td>2</td>
<td>0</td>
<td>0.2372</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>recht</td>
<td>44</td>
<td>214</td>
<td>1</td>
<td>0</td>
<td>0.2359</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gebied</td>
<td>96</td>
<td>183</td>
<td>2</td>
<td>0</td>
<td>0.2563</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dieren</td>
<td>60</td>
<td>97</td>
<td>2</td>
<td>0</td>
<td>0.1599</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hulp</td>
<td>39</td>
<td>115</td>
<td>2</td>
<td>0</td>
<td>0.1599</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stuk</td>
<td>478</td>
<td>42</td>
<td>3</td>
<td>1</td>
<td>0.1599</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>raad</td>
<td>28</td>
<td>130</td>
<td>1</td>
<td>0</td>
<td>0.1592</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>spel</td>
<td>58</td>
<td>97</td>
<td>1</td>
<td>0</td>
<td>0.1586</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dochter</td>
<td>64</td>
<td>93</td>
<td>3</td>
<td>1</td>
<td>0.1583</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>artikel</td>
<td>40</td>
<td>112</td>
<td>2</td>
<td>0</td>
<td>0.1583</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>taak</td>
<td>35</td>
<td>117</td>
<td>1</td>
<td>0</td>
<td>0.1575</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>procent</td>
<td>199</td>
<td>58</td>
<td>2</td>
<td>0</td>
<td>0.1567</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>leden</td>
<td>35</td>
<td>116</td>
<td>2</td>
<td>0</td>
<td>0.1566</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>middel</td>
<td>25</td>
<td>130</td>
<td>2</td>
<td>0</td>
<td>0.1550</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lid</td>
<td>35</td>
<td>114</td>
<td>2</td>
<td>0</td>
<td>0.1550</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>geluid</td>
<td>43</td>
<td>105</td>
<td>1</td>
<td>0</td>
<td>0.1549</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>voet</td>
<td>53</td>
<td>96</td>
<td>3</td>
<td>0</td>
<td>0.1542</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>armen</td>
<td>80</td>
<td>81</td>
<td>1</td>
<td>0</td>
<td>0.1536</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europa</td>
<td>73</td>
<td>84</td>
<td>1</td>
<td>0</td>
<td>0.1536</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


Received October 16, 2006
Revision received February 18, 2007
Accepted April 20, 2007