Optimizing the Multilink model for word translation

Effects of subjective frequency and translation direction

Franka Buytenhuijs (s4356845)
Artificial Intelligence
Radboud University Nijmegen
25-08-2016
Supervisors: Professor Dr. Ton Dijkstra
and Dr. Alexander Wahl
Bachelor thesis

Contents

1 Abstract

In this study, a revised version of the Multilink model (Dijkstra & Rekké, 2012) was used to simulate word recognition and word translation processes, focusing on the effects of frequency and translation direction. Previous studies involving the 2010 version of the Multilink model (e.g., Peacock, 2015) showed that Multilink was able to simulate the empirical data quite well. However, the simulated frequency effect was too small and no translation direction effect occurred, in contrast to the forward facilitation effect in the empirical data by Pruijn (2015).

The simulations indicate that revising the model showed good correlations to the empirical data, and has led to an increase in the frequency, although it is still too small compared to the data by Pruijn (2015). The translation direction effect obtained by Multilink when simulating balanced and unbalanced bilinguals, was also too small compared to the empirical data by Pruijn (2015) but did match empirical studies that found a translation symmetry (e.g. De Groot et al. 1994).

All in all, the revised Multilink model shows great potential in becoming an important all-round psycholinguistic model for further research to word recognition and word translation effects.

2 Introduction

Over the last century, word translation has become more important than ever. As the world transforms into a global village (McLuhan, 1962), translation is required in all important political, economic, social, and scientific activities.

Translation of a written word of one language into a spoken word in a different language is a difficult process that involves several steps, such as word recognition, word retrieval, word translation, and word production in the other language. However, multilinguals are able to complete the process of translation relatively easy and quickly.

The speed and accuracy of word translation depends on the proficiency of the translators in their native and foreign language, but also to considerable extent on the characteristics of the items involved. Important word characteristics are, for example: word frequency (how often does the word occur) and word similarity (how much alike are the words in the different languages). For example, the highly frequent English word HOUSE is translated faster into its similar Dutch translation HUIS than the less frequent word SPICY is translated into the Dutch non-similar translation PITTIG. Another word characteristic that plays an important role is the direction of the translation: that is, whether the translation goes from the native language to the second language or the other way around.

A considerable number of studies have been done to investigate how the human mind is capable of performing these translations. On the basis of these studies, theories have been developed to predict recognition, processing, and translation behavior. However, the large number of processing steps and the complex and interactive effects of word characteristics in the translation task make it difficult to develop a good generalizing theoretical account of all that takes place.

Some earlier models in this field are the Revised Hierarchical Model (Kroll & Stewart, 1994), Bilingual Interactive Activation Plus (Dijkstra & Van Heuven, 2002), and the Multilink model (Dijkstra & Rekké, 2010). In this thesis, I will focus on the frequency effect and the translation direction effect in the translation process as described by the Multilink model. This more recently developed model gives promising results in dealing with the limitations of the earlier models (Peacock, 2015). However, Peacock points out that the frequency effect and the translation direction effect are not yet represented correctly in the model, which makes it necessary to investigate these effects to be able to adapt the model.

First, I will discuss some of the earlier bilingual translation models mentioned above (the Revised Hierarchical Model, the Bilingual Interactive Activation Model, and the Bilingual Interactive Activation Plus Model). Second, I will explain how the Multilink model combines these models into a new theoretical framework. Third, I will clarify the frequency effect and the translation direction effect and the influence they have on each other. Fourth, I will describe the simulations I performed with a revised version of the Multilink model, and show the results of these simulations. Finally, I will discuss the obtained simulation results and draw my conclusions based on these results.

3 Word translation models

3.1 Revised Hierarchical Model (RHM)

The RHM (Kroll & Stewart, 1994; Kroll, Van Hell, Tokowicz & Green, 2010) is a verbal bilingual word processing model that distinguishes two levels: a lexical level, which contains information about word forms, and a conceptual level, which represents meaning (see figure 1).

On the lexical level, a distinction is made between the native language L1 and the second language L2. The L1 is represented as a larger box in the model, because the L1 lexicon of bilinguals contains more words than their L2 lexicon. The conceptual level does not make a distinction between the L1 and L2, but is shared by the two languages.

Both languages are connected to the conceptual system. However, the connection between the L1 and the conceptual level is stronger than the connection between the L2 and the conceptual level. The connection strength has an influence on the processing time, where a stronger connection means less processing time is needed. As a consequence, processing L1 words is faster than processing L2 words.

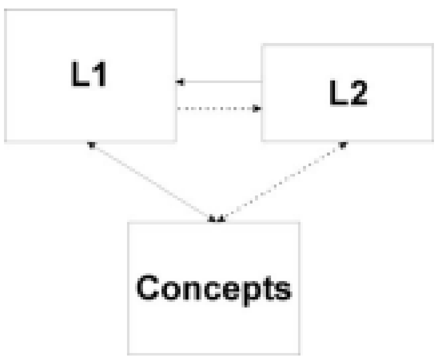

Figure 1 Framework of the RHM model (Kroll & Stewart, 1994)

There are also connections between the two lexical databases. Here, the connection from L2 to L1 is stronger than the connection from L1 to L2, because L2 words are thought to be learned by associating them with the translation from the native language. Therefore, a word in the L2 has a strong connection to the corresponding word in the L1, while the L1 word does not have this connection, or a much weaker one, to the L2 word.

As a result, the RHM contains four different connections: two connections between lexical representations and their conceptual representations, and two connections between the lexical representations. These connections represent two different translation routes (see image 2). The first route proceeds via the semantic level, and is called concept mediation. The input “BIKE” activates the word form bike, and from there first the meaning of ‘bike’ gets activated before the word form of the Dutch translation ‘fiets’ becomes active. The second route directly proceeds to the translation of the word, and is called word association.

According to the RHM, this means that in learners of a new language, translation from L1 to L2 takes place via concept mediation, and translation from L2 to L1 through word association.

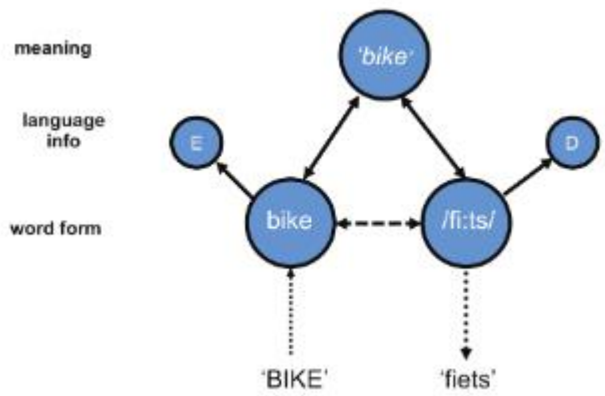

Figure 2 A translation model architecture (Dijkstra & Rekké, 2010)

3.1.1 Limitations of RHM

Brysbaert and Duyck (2010) point out some limitations of the RHM. First, they discuss the lack of evidence for separate lexical representations in each language. They argue that there is evidence against this view. For example, Van Heuven, Dijkstra, and Grainger (1998) found that word candidates compete for recognition both within a language and between languages. These findings for interlingual neighbors perhaps do not reject the hypothesis of separate lexicons, but do make it unlikely.

Second, Brysbaert and Duyck (2010) point out that there is evidence against relatively strong lexical connections from the L2 to the L1. A consequence of this connection is that the translation from L2 to L1 should be faster than the translation from L1 to L2. At the same time, several studies have shown the opposite (Duyck and Warlop, 2009; Pruijn, 2015, in collaboration with Peacock).

Third, implementing the RHM is problematic, because of the simplicity of the model. Although this simplicity seems appealing in the first place, word translation is more complex than described in the RHM. For example, translating words not only involves orthographic but also phonological representations, an aspect that is completely left out of the RHM.

Because of these limitations, Brysbaert and Duyck (2010) suggest moving beyond the RHM and focus on developing the BIA+ model (Dijkstra & Van Heuven, 2002), a bilingual recognition model, into a translation model. I will explain this possibility in a later section of this thesis. Kroll et al. (2010) react to these points of criticism by agreeing that the RHM indeed needs revision after 15 years. However, Kroll claims that the RHM was never meant to be a model of visual word recognition, but that it was designed as a model to describe the asymmetries observed in translation when learning a second language.

3.2 Bilingual interactive activation model (BIA model)

The BIA model (Dijkstra & Van Heuven, 1998) is a bilingual word recognition model that was first developed as an extension of the Interactive Activation (IA) model (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). In the BIA model, the monolingual IA model is extended to the bilingual domain. Thus, the BIA model can be seen as a model that simulates the word recognition phase of word translation in bilinguals.

Like the IA model, the BIA model is a localist-connectionist model, which means that it is made up of nodes that are interconnected (see figure 3). Via these interconnections, nodes can inhibit (restrain another node) or excite each other. Each node corresponds to a symbolic representation (e.g., language, word, letter, letter feature), and each symbol is represented by the activity of a single node. In figure 3, each word corresponds to a node in the lexical level, the same holds for each word concept, and each phonological representation of a word in the other levels of the model.

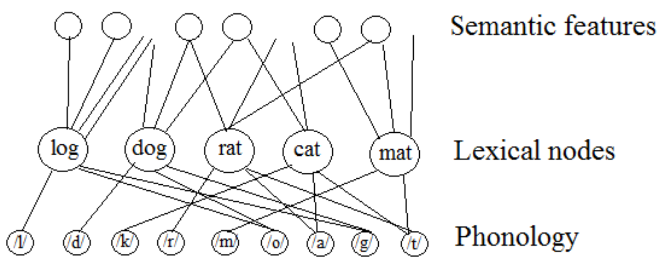

Figure 3 An example of a localist-connectionist model (after Dell & O'Seaghdha, 1991)

Thus, each unit in the BIA model corresponds to a certain feature of the input, like the whole input word, or just a letter or letter feature. The BIA model consists of four different layers representing a different level in lexical access: the feature level, the letter level, the word level, and the language level. This structure of the BIA model is shown in figure 4.

The activation of each unit depends on the strength of the unit it stands for in the input word. For example, when the word BIKE is presented the activation starts at the lowest level (the feature level). When analyzing the input of the letter B, the feature of a vertical bar becomes active, as well as features representing the curves in the bulges of the B. Then, these letter features work together and start activating the letters that they are part of.

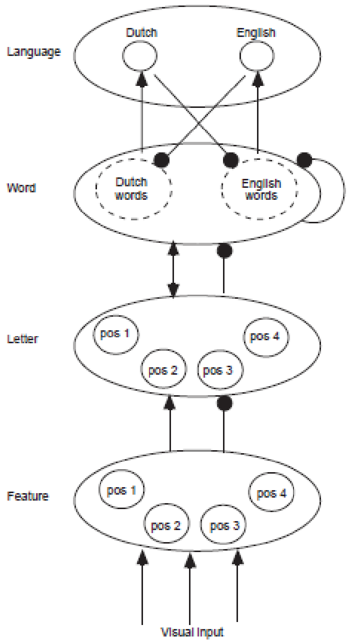

Figure 4 The BIA Model (Dijkstra & Van Heuven, 1998)

However, before any unit becomes recognized, it needs to reach a certain threshold. Once the threshold of a unit is reached, it starts to influence other units. The activation of the units proceeds in a bottom up way, starting at the lowest level. In addition, once a unit is recognized, it also sends feedback activation back downwards. This is the "interactive activation" process of the model. The activation and inhibition in the system eventually leads to one most active word candidate, which is the candidate corresponding to the input word.

One of the biggest advantages of the BIA model over the older RHM model is that it is a computational model instead of a verbal theory. Computational models are more exact than verbal models because they offer quantitative results, whereas verbal models only provide descriptions, and stay based on intuitions. In contrast to verbal models, computational models also make it possible to replicate data in a quantitative rather than a qualitative way.

3.2.1 **Limitations of the BIA model**

The BIA model has certain limitations that should be noted. First, the model does not contain semantic or phonological representations. However, not only orthographic similarity, but also phonological and semantic overlap affect word recognition.

Second, the BIA model is limited to the recognition of words with a fixed length of 4 letters. This places a serious restriction on the words one can translate with the BIA model, and makes it impossible to study effects occurring as a result of word length.

Because of these and other limitations, the BIA model was updated to the BIA+ model (Dijkstra & Van Heuven, 2002).

3.3 Bilingual interactive activation plus model (BIA+ model)

The BIA+ model by Dijkstra and Van Heuven (2002) extended the BIA model with a phonological and a semantic representation (see figure 5). In addition, a task/decision system was included to allow the model to perform different tasks, and in this way study task-dependent effects.

In the identification system, the orthography, semantics, and phonology are all assumed to influence word recognition, and therefore all have an effect on the output that is fed to the task system. The layers in the identification system are fully interconnected, so orthography activates phonology (and vice-versa) at both the lexical and sub-lexical levels. Furthermore, the phonological and orthographic levels both influence the activation of the semantic representations.

3.3.1 **Limitations of the BIA+ model**

The BIA+ model is still fundamentally a word recognition model, and the extension of the recognition model to a translation model leads to several problems pointed out by Dijkstra and Rekké (2010). The biggest problem is that two aspects of the translation task are missed completely. Specifically, in addition to word recognition, meaning representations and word production processes are needed to generate translation output, and these steps are not present in the BIA+ model.

To make up for these shortcomings, Dijkstra and Rekké (2010) developed Multilink, a new computational translation model, that I will discuss in the next section.

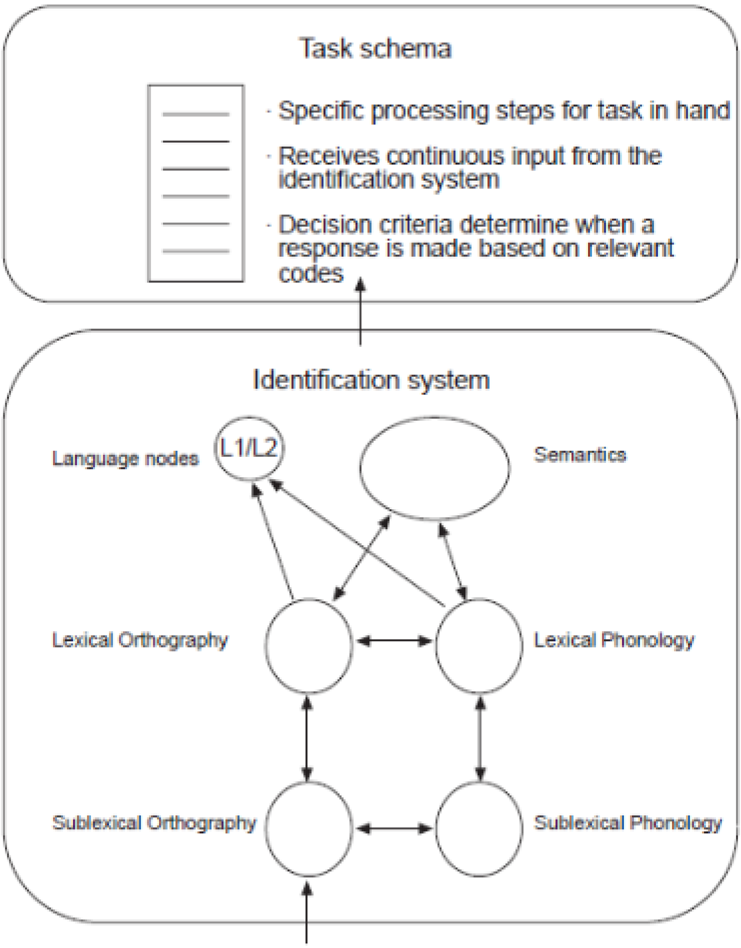

Figure 3 The BIA+ model (Dijkstra & Van Heuven, 2002)

3.4 The 2010 version of Multilink

Multilink is a recently developed bilingual word translation model by Dijkstra and Rekké (2010), that combines important aspects of the RHM and the BIA+ model into a new model. In particular, Multilink shares the notion that the lexicon sizes of the L1 and L2 may differ as in the RHM, but has the same representation of a task/decision system with an identification system as the BIA+ model, shown in figure 6. The different tasks give Multilink the ability to not only explore word translation, but also word recognition, for instance in tasks like lexical decision and language decision. The different decision criteria represent different possibilities to determine when a response is made. For example, when using the decision criterion threshold, the output word corresponds to the word that reaches the activation threshold first.

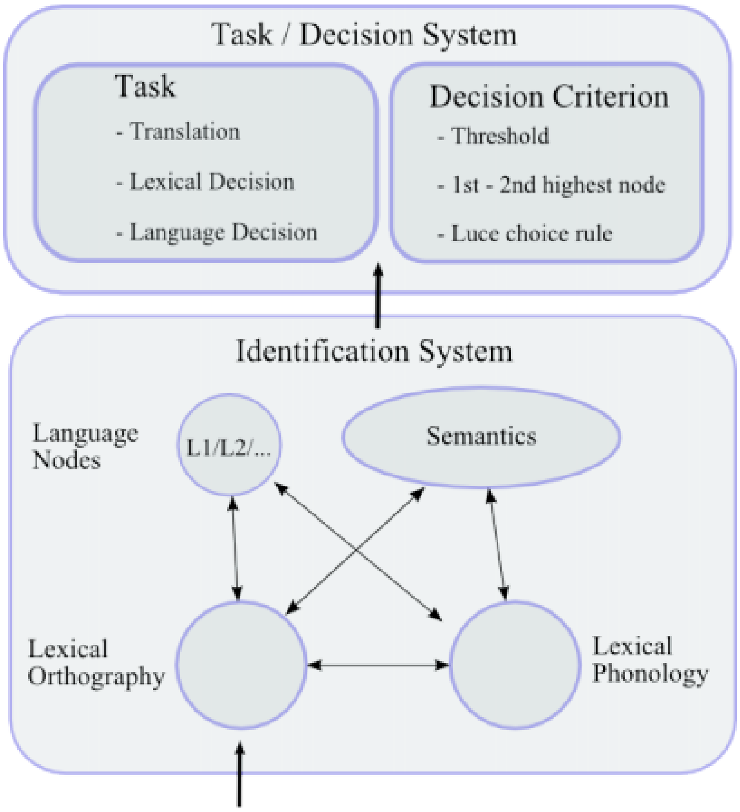

Figure 4 The structure of the Multilink model (Dijkstra & Rekké, 2010)

3.4.1 Implementation of the translation task

In this thesis, I will focus on the translation task that uses a specific activation threshold as decision criterion. Therefore, I will limit my description of the model to this task and decision criterion.

The input to the model is an orthographic letter string, which activates matching graphemic representations in the multilingual lexicon. Next, words with similar lexical-orthographic representations are activated. For example, the input word ANT does not only activate the orthographic representation of the word ANT itself, but also that of form similar words in both languages such as the English word AUNT and the Dutch word TANTE. Currently, this causes some problems in the 2010 version of the model, which I will explain later in this thesis.

The orthographic nodes are connected to language nodes and to semantic nodes, and spread their activation to these other nodes. The semantic nodes and language nodes also give top-down feedback to the corresponding orthographic nodes, the semantic node representing the meaning of ANT, for example, will spread top-down feedback to the orthographic representations of ANT and MIER (the Dutch translation of ANT). Therefore, the top-down activation of the semantic node has an effect whereby words in the other language than the language of the input word (the output language) are able to receive activation. Specifically, since the orthographic nodes in the input and output languages share the same meaning, they both receive top-down activation from this semantic node.

Output is generated when the lexical-phonological representation of a word reaches its activation threshold, but only if this word is from a different language than the original input word. To illustrate this, figure 7 shows the activation spread for the English word FORK, and thereby shows the different steps in the translation of the English word FORK to the Dutch word VORK.

First, when the input word FORK is presented to the model, the word activates orthographic form-similar words like WORK, the Dutch translation VORK, and the word itself (FORK). Second, these orthographic representations activate their semantic representation, which is ‘fork’ in case of the orthographic representation FORK. Third, this semantic representation sends feedback back down to the lexical-phonological representation in the other language that is linked to this meaning. Specifically, that is the Dutch word /vork/ in the case of the semantic node ‘fork’. Finally, the output is the Dutch word in its lexical-orthographic form. In this case the output is “vork”.

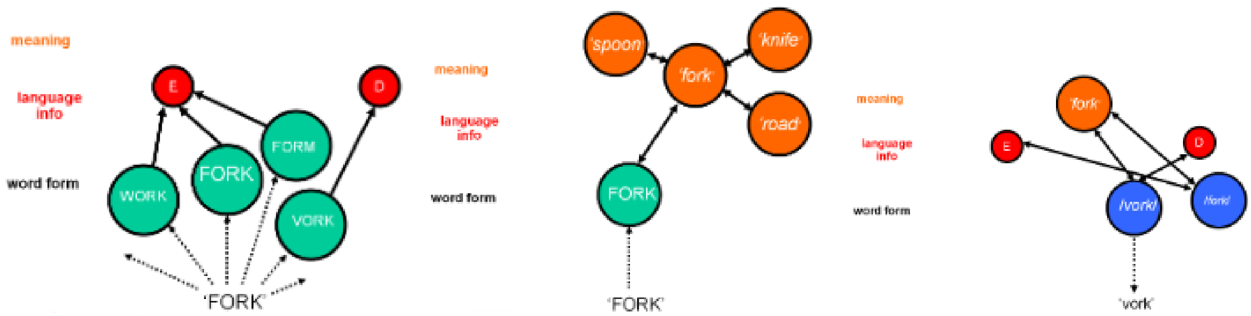

Figure 5 The activation process for the word FORK in Multilink

3.4.2 **How multilink addresses the above-mentioned limitations**

By combining the strengths of different models, Multilink in turn tries to overcome the limitations of these same models. For example, in contrast to the RHM, Multilink is a computational model. Its quantitative results offer a great advantage over the verbal description of the RHM. In addition, in the previous section it was made clear that the architecture of the Multilink model is more complex than the architecture of the BIA+

model. Namely, in contrast to the BIA+ model, Multilink takes into account word production mechanisms in addition to recognition mechanisms. However, in contrast to the BIA+ model, Multilink does not have any sub-lexical layers. Leaving out these sub-lexical layers removes the issue of position-specific encoding entirely, which gives Multilink the freedom to process words of different lengths.

3.4.3 Possible improvements of Multilink 2010

Research on the Multilink model has shown that there are improvements to be made. Peacock (2015) mentions a few important ones, of which I will discuss a few here. First, the cognate facilitation effect, which is the effect that cognates are translated faster than non-cognates, is too strong in Multilink compared to the empirical data. A part of this cognate problem is that for some input words the Multilink model does not find the right translation, but a word in the orthographic neighborhood of the input word. This is for example the case when translating the English word ANT. Instead of finding the Dutch translation MIER, Multilink finds the incorrect translation TANTE. Second, the frequency effect seems to be 5 times too small in Multilink compared to in the empirical data. Third, the translation direction effect as found in the empirical data does not show up in Multilink.

In the next section of this thesis, I will discuss the empirical and simulated effects for these aspects of word translation in more detail.

4 Frequency effect

Word frequency is an important factor in word recognition and word translation. Several studies of the frequency effect, such as Christoffels, de Groot and Kroll (2006), and Pruijn (2015), have shown that participants are faster in the translation of high frequency words than in the translation of low frequency ones in both translation directions. Pruijn (2015) classifies words with a log frequency (base 10) of higher than 1.50 as high frequency words, and words with a log frequency of lower than 1.50 as low frequency words. His results show a clear difference between the RTs in high frequency versus low frequency word translation. These results are presented in figure 13 and figure 14, and will be discussed in more detail in a next section of my thesis.

There is unanimous agreement regarding the existence of the frequency effect and its importance for word recognition. However, the source of the frequency effect is debated by researchers in several studies. I will discuss three theories about how and why the frequency effect occurs.

4.1 Differences in word recognition thresholds

A first possibility is used in the logogen model of Morton (1970), and holds that high frequency words have a lower threshold for recognition than low frequency words. This possibility is shown in figure 8, where a low frequency word (QUAIL) is presented on the left, and a high frequency word (CAT) is presented on the right. Shown here is that the threshold for CAT is lower than the threshold for QUAIL. Therefore, the high frequency word CAT needs less activation to be recognized than the low frequency word QUAIL.

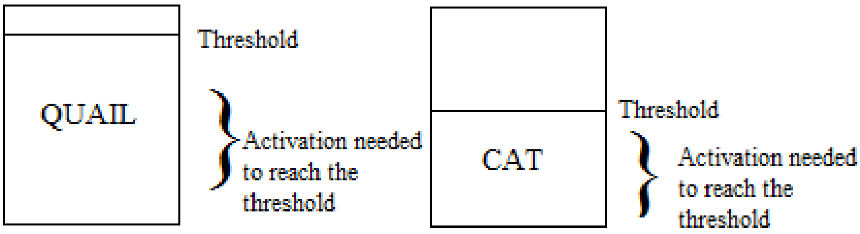

Figure 8 Different word recognition thresholds for high frequency words and low frequency words

4.2 Differences in resting level activation

A second possibility is that the ‘resting level activation’ of high frequency words is higher than the resting level activation of low frequency words. This also means that high frequency words need less additional activation to reach the recognition threshold than low frequency words. The resting level activation approach is used in the BIA+ model (Dijkstra & Van Heuven, 2002), and the Multilink model (Dijkstra & Rekké, 2012). The different resting level activations for high frequency words, and low frequency words are shown in figure 9. Here, the high frequency word CAT, and the low frequency word QUAIL do have the same recognition threshold; however, the level of activation at rest is higher for the high frequency word than for the low frequency one.

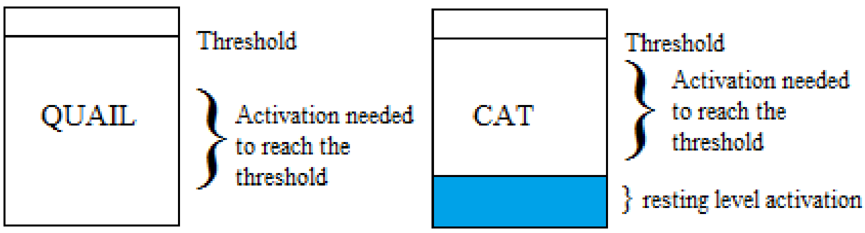

Figure 9 Different resting level activations for high frequency words and low frequency words

4.3 Differences in rate of lexical activation

A third possibility, is that, while words of different frequencies have the same resting level activation and the same threshold, the activation of high frequency words increases faster than low frequency ones. Therefore, the high frequency word reach the recognition threshold faster than the low frequency word. This possibility is shown in figure 10, and is used in the Cohort model (Marlsen-Wilson, 1987).

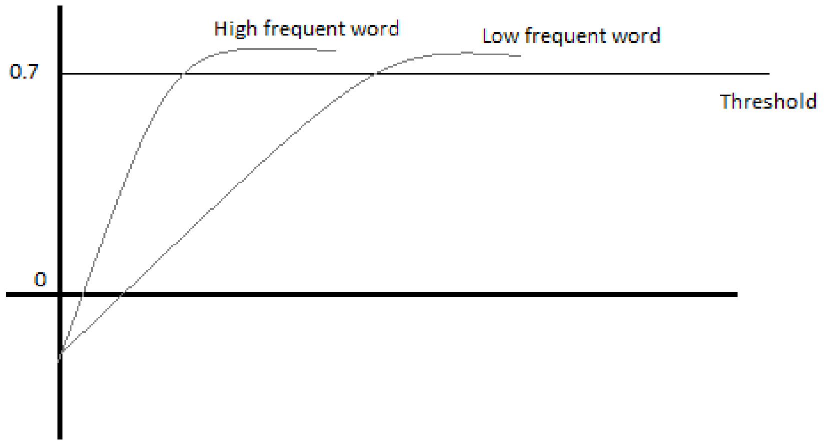

Figure 10 Different in rate of lexical activation for high frequency words and low frequency words

4.4 Frequency effect in Multilink 2010

In the Multilink model, the frequency effect occurs as a result of different resting level activations. Before a word is recognized it needs to reach a certain threshold. Words with a higher resting level activation are already closer to this threshold in their resting state. Therefore, when a lexical item is presented, words with a higher resting level activation are recognized faster than words with a lower resting level activation.

4.4.1 Resting level activation in Multilink 2010

In Multilink 2010, a ranking system is used for the calculation of the resting level activation. The idea to use this ranking implementation comes from the Interactive Activation model (McClelland & Rumelhart, 1988). This ranking system is implemented in Java via the formula in equation 1.

$$\text{MIN\_REST} + \text{RANK} * (\text{Math.abs}(\text{MIN\_REST} - \text{MAX\_REST}) \,/\, \text{MAX\_RANK}) \tag{1}$$

In the model, the MIN_REST variable represents the lower bound for the resting level activation, and is set to -0.05. The MAX_REST variable represents the upper bound for the resting level activation, and is set to 0. The Java function *Math.abs()* takes the absolute value of the difference between the MIN_REST and the MAX_REST.

RANK is the rank of the current word for which we want to determine the resting level activation. The rank is derived from the frequency of the word, where the higher the frequency of the word, the higher the rank. Words with the same frequency get the same rank. As a consequence, the MAX_RANK is rank of the highest frequency word in the lexicon.

5 Cognates

Cognates are words with the same meaning in two languages which show a form overlap (for example, the Dutch word TOMAAT and the English word TOMATO). As a result of this overlap, studies (e.g. Pruijn, 2015) have shown that cognates are translated faster than non-cognates. The results from Pruijn (2015) are shown in figures 13 and 14 in a next section of this thesis, where the graphs show the RTs for a translation task. For both the translation from English to Dutch and from Dutch to English, there is a clear difference between the RTs for cognates and non-cognates, for the high frequency words as well as for the low frequency words. As can be seen in these graphs, the cognates are translated significantly faster than the non-cognates.

5.1 Cognate facilitation in Multilink 2010

The cognate facilitation effect is also implemented in the Multilink model. To take the cognate effect into account, we first have to determine if a word and its translation are cognates. In Multilink, this is calculated using the Levenshtein distance measure (Levenshtein, 1966). This measure calculates the distance between two letter-symbol sequences by making use of three possible operations: insertion, deletion, and substitution. The use of one of these operation adds one up to the total Levenshtein distance each time it is used.

The insertion can be used to add a symbol, the deletion to remove a symbol, and the substitution to remove one symbol and replace it by another symbol. Formally, this is calculated according to equation 2. For the example I gave earlier (the Dutch TOMAAT and the English TOMATO), this means that there is a Levenshtein distance of two. This is calculated by using one deletion (one of the A's) and one insertion (the final O), when transforming the Dutch word to the English word. Because there are two operations needed, the total Levenshtein distance is 2.

$$\mathrm{LD}(\mathrm{i},\mathrm{j}) = \min\left\{\begin{matrix} \mathrm{LD}(\mathrm{i}-1,\mathrm{j})+1 \\ \mathrm{LD}(\mathrm{i},\mathrm{j}-1)+1 \\ \mathrm{LD}(\mathrm{i}-1,\mathrm{j}-1)+\begin{cases} 1 \text{ if input(i)} \neq \text{destination(j)} \\ 0 \text{ if input(i)} = \text{destination(j)} \end{cases} \end{matrix}\right\} \quad (2)$$

Using the Levenshtein distance has two advantages: first, it is possible to activate words with different lengths simultaneously, by normalizing the score afterwards. Second, the Levenshtein distance is not position specific; therefore, it can handle words of different lengths. The formula used in Multilink to calculate the normalized Levenshtein distance between two words is given in equation 3 (Dijkstra and Rekké, 2012).

$$\text{Score} = \left(\frac{\max(|\text{input}|,|\text{destination}|) - \text{LD}(\text{input}, \text{destination})}{\max(|\text{input}|,|\text{destination}|)}\right) \qquad (3)$$

The second step in the cognate facilitation effect is the actual facilitation of the found cognates. This is done with the IO-Multiplier parameter, which is an arbitrary chosen value, set at 0.1755 in the current version of the model. If the normalized Levenshtein distance is larger than 0.5, the IO-multiplier is multiplied by the square of the Levenshtein distance. This multiplication gives the final score, which is the boost for the word. If the Levenshtein distance is smaller than 0.5, which means that the input word and the destination word are non-cognates, the boost is set equal to 0, see equation 4.

$$\text{score} = \left\{\begin{matrix} \text{if } \left(\frac{\text{MAX}_\text{L} - \text{LD}}{\text{MAX}_\text{L}}\right) > 0.5 \ \ then & \text{IO}_\text{multiplier} * \left(\frac{\text{MAX}_\text{L} - \text{LD}}{\text{MAX}_\text{L}}\right)^2 \\ \text{Otherwise} & 0 \end{matrix}\right\} \qquad (4)$$

Note that this implementation has a disadvantage whereby a word can easily be translated into a word that is not the correct translation, but which has a small Levenshtein distance with respect to the input word. This is, for example, the case when translating the Dutch word ANT. The expected output in this case is the correct English translation MIER. However, the model outputs the word TANTE as the translation for ANT, because the Levenshtein distance between the word ANT and the word TANTE is just two, and the Levenshtein distance between ANT and AUNT (the English translation of TANTE) is only one. As a consequence of their similarity, ANT-TANTE, and ANT-AUNT are both recognized as cognates, which gives these words a cognate boost in activation. The shared semantic node of TANTE and AUNT gets more activation than the semantic node of ANT, and as a consequence the node of TANTE and AUNT will reach the threshold faster than the semantic node of ANT. The semantic node of AUNT and TANTE spreads its activation to the phonological nodes /aunt/ and /tante/. However, because /aunt/ is not a word in the right output language (the other language than the input language, so in this case Dutch), the final output word becomes TANTE. The activation of the different words, after presentation of the input word ANT, is shown in figure 11.

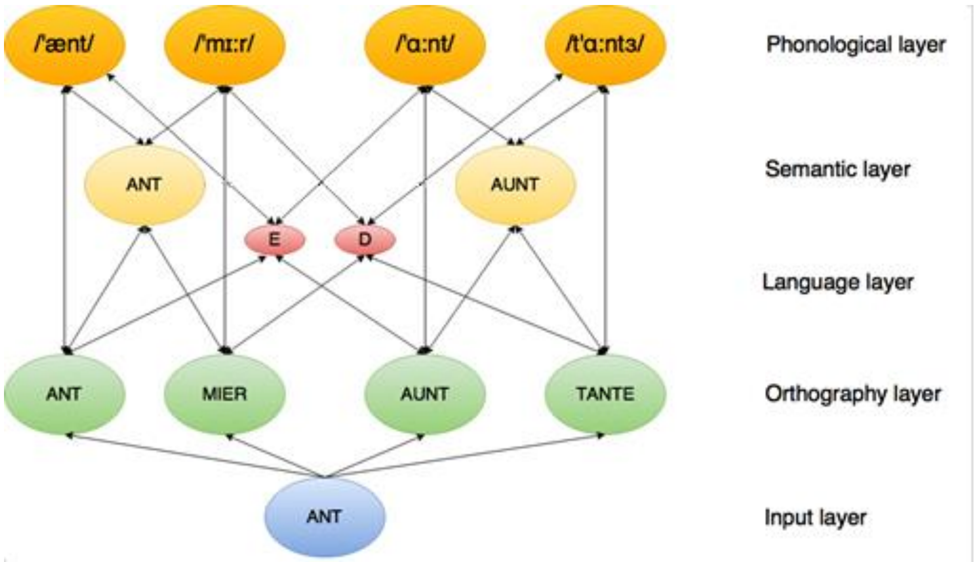

Figure 11 Activation spreading after the input word ANT is presented in Multilink

6 Translation asymmetry

6.1 Unbalanced bilingualism

In the research previously completed using the Multilink model, the frequencies for the native language (L1) and the second language (L2) words were taken from the frequency counts of the Centre for Lexical Information database (CELEX, Burnage, Baayen, Piepenbrock & van Rijn, 1990). The outcomes of the simulations with the Multilink model therefore reflect the lexical processing of a balanced bilingual (i.e., English word frequencies were from native speaker discourse), while most bilinguals are unbalanced bilinguals. Unbalanced bilingualism implies there is a so-called ‘strong’ language and a so-called ‘weak’ language. The ‘strong’ language is the native language, which is used more often. The ‘weak’ language is a second language.

6.2 Translation direction effect

Several researchers like Kroll and Stewart (1994), Christoffels et al. (2006), and Pruijn (2015) have found that unbalanced bilingualism leads to a translation direction effect: that is, that translation in one direction goes faster than the translation in the other direction.

Translations can proceed from the native language to the foreign language (forward translation, from L1 to L2), or in the opposite direction (backward translation, from L2 to L1). Because the participants in the several empirical studies are unbalanced bilinguals, we expect to find a difference between forward and backward translation. Specifically, because the English words are used less often than the Dutch words we expect that the recognition and retrieval of the English words is slower than the recognition and retrieval of the Dutch words. In turn, we expect that these recognition/retrieval differences have an effect on the overall time it takes to translate words. The question is in which direction the translation is facilitated, and to what extent this facilitation occurs. Pruijn (2015) discusses three possible outcomes for studies on translation facilitation: backward facilitation, forward facilitation and translation symmetry, all of which have been found in different empirical studies (see table 1).

	Forward Translation	**Backward Translation**	**Overall found effect**
Kroll & Stewart (1994) Experiment 3 Participants: Students	1269 (51.3%)	1140 (63.5%)	Backward facilitation
Christoffels et al. (2006) Experiment 1 Participants: Students	912 (90.4%)	978 (89.0%)	Forward facilitation
De Groot et al. (1994) Experiment 1 Participants: Students	1307 (80.4%)	1271 (86.3%)	Null

Table 1 Found reaction times (in milliseconds) in three earlier translation studies, and the translation direction effect they found.

6.2.1 Backward facilitation

Kroll and Stewart (1994) completed an empirical translation study to test the Revised Hierarchical Model, where they found faster translations from L2 to L1.

They explained this backward facilitation effect by arguing that forward translation follows the slower concept mediation route than the lexically mediated backwards translation (see figure 2), and therefore takes more conceptual memory. In addition, a more recent study by Kroll, Michael, Tokowicz, and Dufour (2002) also found a significant backwards facilitation effect.

6.2.2 **Forward facilitation**

The second possibility discussed by Pruijn (2015) is a forward facilitation effect, where the translation from L1 to L2 is faster than the translation from L2 to L1. Christoffels et al. (2006) found this effect in their empirical study. They created eight word categories by manipulating three variables: cognate status (cognates or non-cognates), frequency (high frequency or low frequency) and translation direction (forward translation [Dutch to English] or backwards translation [English to Dutch]). Of interest here was that the frequency effect and cognate effect they found did not significantly interact with the translation direction. In other words, the facilitation they found in this study occurred independently of the other factors they manipulated.

Pruijn also found a forward facilitation effect in his empirical study. The stimuli he used varied in terms of the same categories as the stimuli used by Christoffels et al. (2006). His results are shown in table 2.

		English-Dutch	**Dutch-English**	**Language Effect**
Cognates	**HF**	737 (98.6%)	726 (97.0%)	-11
	LF	826 (95.2%)	808 (93.1%)	-18
Non-cognates	**HF**	847 (96.9%)	800 (95.9%)	-47
	LF	981 (88.7%)	917 (90.2%)	-64
Average		848	813	-35*

Table 2 Results on translation direction effect found by Pruijn (2015)

6.2.3 Translation symmetry

The last possibility discussed by Pruijn (2015) is translation symmetry. In this case there would be no facilitation in either direction, which means that word translation from L1 to L2 is performed as fast as translation from L2 to L1. This possibility has also been observed empirically, in a study by De Groot, Dannenburg, and Van Hell (1994).

6.3 Conclusions on translation direction effects

In the previous section, I discussed three different possible translation direction effects, which all have found empirical evidence. While the RHM and the research conducted by Kroll et al. (1994) point to a backward facilitation effect, the more recent results of Christoffels et al. (2006) and Pruijn (2015) show a forward facilitation effect. In contrast, De Groot et al. (1994) did not find a facilitation effect at all.

I will compare results from Multilink with the results found in the recent studies by Christoffels et al. (2006) and Pruijn (2015). I will describe their experiments in a later section of my thesis.

6.4 Translation asymmetry in Multilink 2010

In the simulations conducted with Multilink 2010 by Peacock (2015), no translation asymmetry was found, as can be seen in the histogram in figure 12. In the histogram, the average cycle times from the model are shown in the forward and backward translation condition.

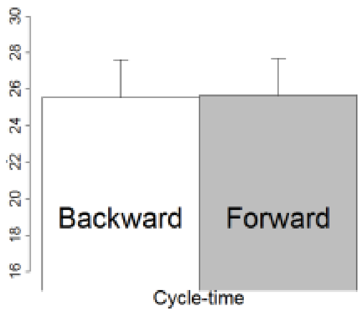

Figure 12 Histogram comparing mean cycle-time for the backward and forward conditions. (Peacock, 2015)

6.5 Modeling translation asymmetry by manipulating frequency

As discussed above, the Multilink 2010 output data does not show a forward facilitation effect, as in the data by Pruijn (2015). This may be because the participants of the empirical study by Pruijn (2015) were all unbalanced bilinguals. However, the stimuli used in previous research with the Multilink model all assume use by balanced bilinguals, as the word frequencies used in the lexicon are, for the L1 as well as the L2, token frequencies from the CELEX database (Burnage et al., 1990). That is, for both languages, word frequencies are used that represent the word frequencies of native speakers. If direction asymmetry is caused because the participants were unbalanced bilinguals, a translation direction effect might occur if we change the word frequencies used in the simulations.

Changing the L2 frequencies should make them more in line with the individual characteristics of the participants in the experimental studies. In their research with the BIA model, Dijkstra, Van Heuven, and Grainger (1998) argued that the simulation results fit better to the empirical data when they used so called ‘subjective’ frequencies: a reduced frequency range for the words in the L2 (calculated, for example, by dividing the native-speaker frequencies by a constant). This causes the maximum resting level activations of the L2 words to drop below that of the L1 words. Therefore, the L2 words are activated somewhat more slowly than the L1 words.

The suggestion of dividing the frequencies by a constant, was not based on empirical evidence, and therefore Duyck, Van der Elst, Desmet, and Hartsuiker (2008), and Cop, Keuleers, Drieghe, and Duyck (2015) did studies on the size of the frequency effect in L1 and L2 for unbalanced bilinguals in word recognition. Duyck et al. (2008) focused on the differences between the frequency effect in L1 and in L2. In two experiments they found that bilinguals show a significantly higher frequency effect in their L2 than in their L1, and that this language x frequency interaction was not due to confounding English stimuli used in the experiment. Cop et al. (2015) did an experiment where they used a natural reading task to study the frequency effect in bilinguals. They used eye-tracking in order to analyze the fixation durations of the participants on the words in the novel they had to read. Their results also showed a larger frequency effect in the L2 than in the L1, and they found that this difference was indeed due to the proficiency of the participants.

The lower L2 resting level activations in the BIA+ model would predict a larger frequency effect in the L2, because the L2 words will, in comparison to the L1 of the same corpus frequency, generally have lower resting level activations. Therefore, these findings support the theory that the L2 words should have relative lower frequencies than the L1 words, giving them a lower resting level activation in the model as argued by Dijkstra et al. (1998) in the BIA model.

An unanswered question that remains is in which respects the unbalanced L2 lexicon is different from the L1 lexicon. To do justice to the lower subjective frequency of the L2, Dijkstra et al. (1998) proposes to divide the L1 frequencies by a constant, such as 4 or 10. This is a linear transformation of the lexicon, which means that all words, high frequency and low frequency, are treated equally, and that the distribution of the frequencies in the lexicon will not change. However, it is questionable if this is a correct way to model the L2 frequencies, and if it would not be better to apply a non-linear transformation instead.

A linear transformation suggests that bilinguals use all the words in their L2 lexicon an equal amount less than native speaker do. However, it might be that low frequency L2 words are used less often than that or even never, because they are not known by the bilinguals. Another option is that a linear transformation might overestimate the use of high frequency words in bilinguals. If this is the case, a nonlinear transformation would be better in order to fit the empirical data.

A non-linear transformation, like the square or the square root, does change the frequency distributions in the lexicon. For example, applying the square root will ‘punish’ high frequency word more than low frequency words. If we have two words: A and B, where A has a frequency of 4 and B has a frequency of 16. Applying the square root to these words will make word A just 2 times less frequent, while word B becomes 4 times less frequent. However, it is not clear if a non-linear transformation is better than a linear transformation.

7 Methodology

In this section of my thesis I will describe my research on the basis of my research questions. Thereby, I will explain how I obtained the answers to my research questions by describing the revised version of the Multilink model and the experimental data I used.

First, I will explain my research questions in section 7.1. Second, I will go into further detail to the study by Pruijn (2015), from which I used the empirical data. Lastly, I will explain the revisions made to the Multilink 2010 model to obtain a new version of the model (Multilink 2016).

7.1 Research questions

In the previous sections, we discussed different implementation issues in the Multilink model. Specifically, as stated by Peacock (2015), the frequency effect is too small in the Multilink output data, while the cognate effect is too big, and the translation direction effect does not show up as in the empirical data by Pruijn (2015). In my simulations with the Multilink model, I will focus on two effects: the frequency effect and the translation direction effect.

In particular, the goal of this thesis is to further optimize these effects in the Multilink model, with the following two research questions:

1. *What modifications to the resting level activation can be done to increase the frequency effect in the model?*

2. *Can the simulation of unbalanced bilinguals increase the forward translation effect in the Multilink model?*

These questions will be answered by comparing the word translation data obtained by Pruijn (2015) with the data from the 2016 version of the Multilink model. These will be discussed in the next sections.

7.2 Experimental data

To evaluate the performance of the model, the empirical data collected by Pruijn (2015) will be used to compare the empirical RTs with the cycle times from the model. In the study by Pruijn (2015), vocal translation latencies were collected in eight different categories, by crossing three binary categories: cognate status (cognate or non-cognate), frequency (high frequency or low frequency), and translation direction (forward translation [Dutch to English] or backward translation [English to Dutch]).

He collected the latency times of 256 items (32 items in each category) from 42 bilingual participants. From this collected data, the data of five participants and three words were removed, because of their poor performance. From the remaining data, the average of each item was calculated, by averaging the individual RTs per word. These average RTs per word will be used as comparison for the Multilink output data.

The results from Pruijn (2015) are presented in figure 13 and figure 14, and show that high frequency words are translated faster than low frequency words, that cognates are translated faster than non-cognates, and that translation from L1 to L2 is faster than translation from L2 to L1. These results lay in line with the results found by Christoffels et al. (2006), who used the same eight word categories in their research.

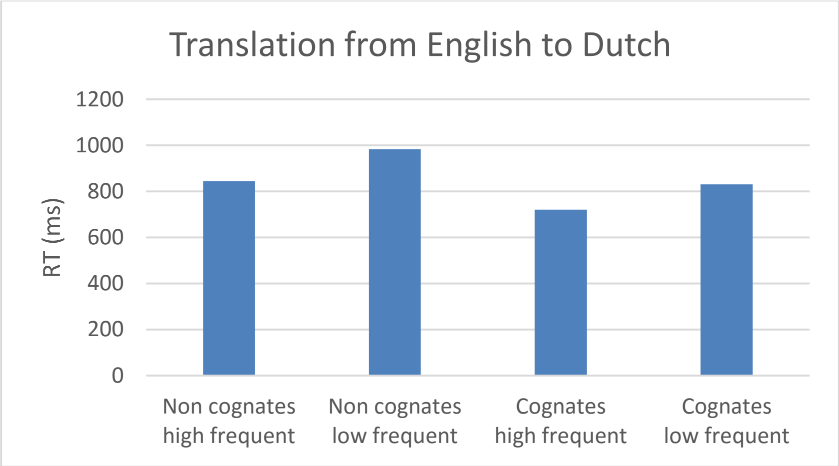

Figure 13 Results found by Pruijn (2015) for translation from English to Dutch

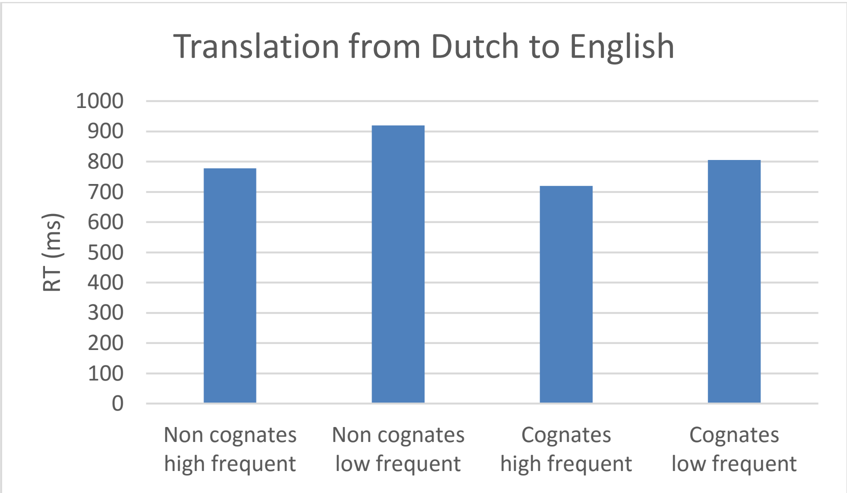

Figure 14 Results found by Pruijn (2015) for translation from Dutch to English

7.3 The 2016 version of the Multilink model

For the simulations with the Multilink model, I use a revised Multilink model by Rekké, Al-Jibouri, Buytenhuijs, De Korte, and Van Halem (2016), developed in collaboration with Dijkstra and Wahl. We adjusted the lexicon, score measure, frequency ranking, resting level activation, and added a phonological representation. I will discuss these changes in more detail in the section below.

7.3.1 Phonology

In the 2010 version of the Multilink model, no phonological representation was included. Therefore, the P nodes (phonological nodes) of the two languages were also represented as O nodes (orthographic nodes). In the new version of the model, we replaced these O nodes by P nodes, by replacing the orthographic word representations by phonological word representations.

7.3.2 Lexicon

We conducted simulations using a lexicon that was adapted with respect to three different aspects: We changed the incorporated frequencies, we removed certain words, and we replaced the original orthographic output representation with a more realistic phonological representation. The adjusted lexicon can be found in appendix A.

-Frequencies

The new frequencies used in the Multilink simulations were gathered from SUBTLEX-NL (Keuleers, Brysbaert, & New 2010) and SUBTLEX-USA (Brysbaert & New, 2009), instead of the previously used CELEX (Burnage et al., 1990) database. SUBTLEX is a database that contains word frequencies based on the subtitles of films and television shows. In comparison to the CELEX database, SUBTLEX-NL is able to explain up to 10% more variance in RTs and accuracies. Because of these higher explained variance rates, Keuleers et al. (2010) argue that the use of SUBTLEX for experiments should be preferred above CELEX.

The SUBTLEX frequencies represent the ‘objective’ frequencies of usage by native language speakers: They represent the average monolingual’s frequency distribution of the language.

-Words

We excluded three categories of problematic words from the lexical database. The first category are verbs, because they are difficult to use in modeling the translation task. Verbs can occur in different forms, because we conjugate them when using them in everyday language. This makes it more difficult to assess which frequency needs to be used in a simulation: the lemma frequency, summing up the frequency of all conjugations, or the word form frequency. Extra theoretical steps would be needed to model the different forms of a verb, and their interaction. In the end, we decided to leave them out of current simulations.

The second group of words we left out are identical cognate pairs, like FREQUENT - FREQUENT. This word is both an English and a Dutch word which share the same concept. These identical cognates must follow a different translation route than non-cognates. Dijkstra et al. (2010) found that identical cognates are processed faster than nearly identical cognates, like MELOEN and MELON. They found a discontinuously large facilitation effect for the identical cognates in comparison to the nearly identical cognates. Because of the special way in which identical cognates may be represented, it may be better to leave them out of simulations until more is known about their representation and processing.

The last group of items we removed are pronouns and determiners. Psycholinguists have proposed that such function words may not be recognized in the same way as content words (like nouns and verbs). Furthermore, a big problem with these words is that most of them are of very high frequency. In the model, such high frequency items may interfere with the processing of other similar words.

-Phonological representation

In the revised model (Multilink 2016), a phonological representation is added. Therefore, a phonological representation in the lexicon is required to produce the right phonological output. We took the phonological representations from the CELEX database (Burnage et al., 1990), which offers a phonological representation in ASCII code.

7.3.3 Score measure

As I explained earlier in my thesis, the Levenshtein distance is used to calculate the similarity of two words. When the normalized Levenshtein distance between 2 words is bigger than 0.5, words get a certain ‘boost’. This boost ensures that cognates are recognized faster than non-cognates, and is calculated by a score function, which is illustrated in equation 4.

$$\text{score} = \left\{\begin{matrix} \text{if } \left(\frac{\text{MAX}_\text{L} - \text{LD}}{\text{MAX}_\text{L}}\right) > 0.5 \;\; then & \text{IO}_\text{multiplier} * \left(\frac{\text{MAX}_\text{L} - \text{LD}}{\text{MAX}_\text{L}}\right)^2 \\ \text{Otherwise} & 0 \end{matrix}\right\} \quad (4)$$

Note that this formula has some disadvantages. First, words only get a boost when the normalized Levenshtein distance to the input word is bigger than 0.5. This means that words with a normalized Levenshtein distance of 0.49 and words with a normalized Levenshtein distance of 0 are treated equally by the model. Concretely, this means that the words bike – book are treated the same as the words bike – zero.

Second, Pruijn (2015) used a different definition of cognates than this definition in Multilink. In his study, words with a Levenshtein distance of 3 or less were classified as cognates. Because we are using his empirical data as comparison for the Multilink outputs, this is a problem. As a consequence, not all the cognate stimuli used by Pruijn (2015) were recognized as cognates by Multilink, and conversely some of the non-cognate stimuli used by Pruijn (2015) were recognized by the model as cognates.

Because of these two problems, we decided to get rid of the condition that the normalized Levenshtein distance must be bigger than 0.5, and instead calculate a score for all the words, to also take effects occurring through small form overlap in account.

Previous research (Peacock, 2015) found that the cognate effect was too big when squaring the score measure, since an orthographic similar word could get recognized, even when it was not the correct translation of the input word. This was for example the case when translating the English word ANT. Instead of finding MIER as a translation, Multilink found TANTE as a translation, because of the similarity between ANT and TANTE (see above for discussion).

To overcome this problem, we changed the formula: instead of taking the square of the score measure, we now take its cube. Because the score measure is a value between zero and 1, cubing the score measure gives a smaller value than squaring the score measure. Therefore, similarity has a less dominant role in the translation process in Multilink this way.

These two changes together give us the new score measure, as illustrated in equation 5

$$\text{score} = \text{IO}_{\text{Multiplier}} * \left(\frac{\text{MAX}_{\text{L}} - \text{LD}}{\text{MAX}_{\text{L}}}\right)^3 \qquad (5)$$

7.3.4 Frequency ranking

In the previous version of the model, a ranking method was used to calculate the resting level activations. The disadvantage of the ranking system is that variation in frequency differences between two consecutive words have no effect on the score. As an example, when we use a lexicon with three words, and the frequencies of these words are 40, 41 and 600, the ranking of these three words will be the same if they would have had the frequencies 40, 599, and 600. Because the words get the same ranking in these two situations, the resting level activations across these situations will be equal as well, even though the frequency distributions are quite different.

To allow the frequency differences between words in the lexicon to influence the resting level activations, we decided to change the ranking system into a formula where we use the logarithm of the frequency. Tests with the base 10 logarithm as well as the natural logarithm showed that the base of the logarithm did not have an influence on the distribution, so we arbitrarily decided to use the base 10 logarithm. To make sure the base 10 logarithm of the frequency of all items is above 0, we changed the occurrences per million to occurrences per billion by multiplying the SUBTLEX frequencies by 1000 before taking the logarithm.

These changes resulted in the formula for calculation of the resting level activations seen in equation 6.

$$(\mathrm{MIN\_REST} + \mathrm{Math.10log(OPB)} * (\mathrm{Math.abs(MIN\_REST - MAX\_REST)} / \mathrm{Math.10log(MAX\_OPB)}) \quad (6)$$

Where OPB stands for occurrences per billion.

Although this method is empirically more plausible (since this way the real frequency differences between two words is taken in account), some of the correlations between the Multilink results and the empirical data by Pruijn (2015) actually drop. This is especially the case for the low frequency cognates (see figure 15), on which I will focus more in a later section of my thesis. Because of this problem, the issue of the resting level activation needs further research.

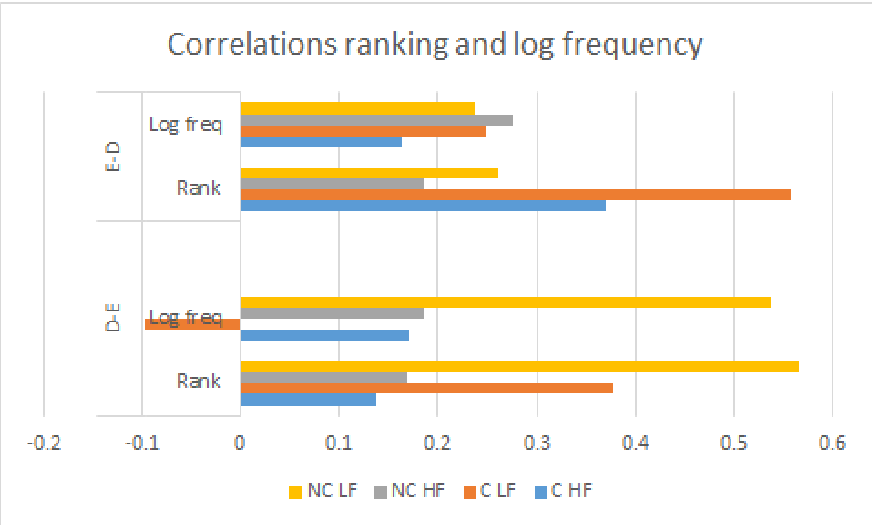

Figure 15 Correlations of Multilink outputs and Pruijn (2015) for ranking system and use of log frequency

7.3.5 Resting level activation

The resting level activation is a value each word has when at rest. The higher the frequency of a word, the higher the resting level activation of that word. In the earlier version of the Multilink model, the resting level activation was a value between 0 and -0.05. The old formula for the resting level activation was shown in equation 1, and is presented again below.

MIN_REST + RANK * (Math.abs(MIN_REST - MAX_REST) / MAX_RANK)   (1)

The value -.05 is a value taken from earlier models, such as the RHM. However, using parameter fitting, we found that a resting level activation with a minimum value between -.15 and -.25 results in a much better fit to the empirical data. Therefore, we decided to change the minimum resting level activation to -.20. This alteration should increase the frequency effect in the Multilink model, because the differences between low frequency words and high frequency words increases with a larger range for the resting level activations. When we include the changes we made to the ranking system, the formula changes to that in the previous shown equation 6.

Additionally, we made a second change in the calculation of the resting level activation. In the equation 1, the resting level of a word was dependent on the rank of the most frequent word. However, by changing the ranking system to the new system, whereby we use the logarithm of the frequency to calculate its resting level activation, the resting level activation has become directly dependent on the frequency of the highest frequency word in the lexicon. As a consequence, when we add a new word to the lexicon with a higher frequency than all the current words in the lexicon, the resting level activations of all words in the lexicon change. In the earlier formula, this dependency was less dramatic, because adding a new word increased the MAX_RANK by a maximum of 1. However, in using the occurrences per billion, the changes made to the resting level activations can be much larger.

For example, when we have a lexicon with 1000 words that all have a different frequency, the MAX_RANK in the 2010 version of the model is always 1000, independent on the individual frequencies of the words. Adding a word to the lexicon can increase this MAX_RANK by a maximum of 1. However, in the 2016 version of the model the MAX_OPB represents the logarithm of the frequency of the highest frequency word in the lexicon. Therefore, when the highest frequency word is 50 (in which case the MAX_OPB will be about 1.70), all the resting level activations will change more dramatically when adding a new word with a frequency of, for example, 50000 (in which case the MAX_OPB will be almost 4.69).

To avoid the dependency on the frequency of words in the lexicon, we decided to change the value of MAX_OPB to a constant: the value of the base 10 logarithm of the most frequent word in the Dutch and English SUBTLEX databases, which is the English word *the* with a frequency of 46692368 per billion. Therefore, we changed the value of the MAX_OPB to the base 10 logarithm of 46692368.

8 Simulation 1: Word recognition in monolinguals

Word recognition is the first step in word translation, and therefore it is important to see how a translation model performs on this task. Since early problems will affect the performance of the model in later phases, problem detection in the early steps of the model is very important. The most basic situation is word recognition in monolinguals. In this simulation, I will compare Multilink 2016 to the IA model (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), the monolingual predecessor of the BIA model (Dijkstra & Van Heuven, 1998). Through this model-to-model comparison, we can see on which aspects of word recognition Multilink performs better than the earlier models, and on which aspects the earlier models perform better. This way, we can see the value of Multilink to the field of linguistic models, and also find possible improvements on the basis of the implementation of the earlier IA model.

8.1 Method

For this simulation, I am using the most recent online implementation of the IA model (http://www.psychology.nottingham.ac.uk/staff/wvh/index.html, 2015) by Walter van Heuven in jIAM, and the recognition task implemented in the adapted Multilink model. I am using the already implemented lexicon in jIAM for the IA model, which I have adapted to a format that is also usable in the Multilink model. From this lexicon I removed all the cognates, because cognates add complications to the simulations that we do not wish to consider here. The used input words can be found in appendix B.

I compared the outcomes of the runs of the IA model, and the Multilink model with data from the British Lexicon Project (BLP; Keuleers, Lacey, Rastle, & Brysbaert, 2012), and in particular examined the frequency effects occurring in the empirical data versus the models.

For word frequencies, I used the same definitions for high frequency and low frequency as Pruijn (2015), which means that all words with a base 10 logarithm of the frequency below 1.50 are classified as low frequency, and all words with a base 10 logarithm of the frequency above 1.50 are classified as high frequency.

In addition, I converted the Multilink cycle times to RTs in milliseconds in order to compare model times with empirical data. I multiplied the cycles times with a value obtained by the division of the average of all the RTs by the BLP divided by the average over all Multilink cycle times, which is also shown in equation 8. The cycle times for the IA model were scaled in a similar way.

$$\mathrm{RT(i)} = \mathrm{Cycle\ time(i)} * \frac{\mu\ \mathrm{(RTs\ of\ the\ emperical\ data)}}{\mu\ \mathrm{(Cycle\ times\ of\ the\ model)}} \qquad (8)$$

8.2 Results

The correlation of the Multilink 2016 model with the BLP turned out to be higher than the correlation of the IA model with the BLP (see figure 16). However, when we plot the RTs against the logarithm of the frequency, the IA seems to show more clearly a curve similar to that in the empirical data than Multilink (figure 17).

For the very high frequency words, the Multilink latencies are too low compared to the BLP, while the IA latencies for very high frequent words are closer to those of the BLP data. As a consequence, the frequency effect is too big in the Multilink model, as can be seen in figure 18, where the average reaction time of the low frequency words is subtracted from the average reaction time of the high frequency words.

A second observation, which can be seen in image 17, is the spreading of data points. The BLP shows a large amount of variance, while the variance is much smaller in the Multilink data. The IA model shows more variance than the Multilink model, but still less than the empirical data.

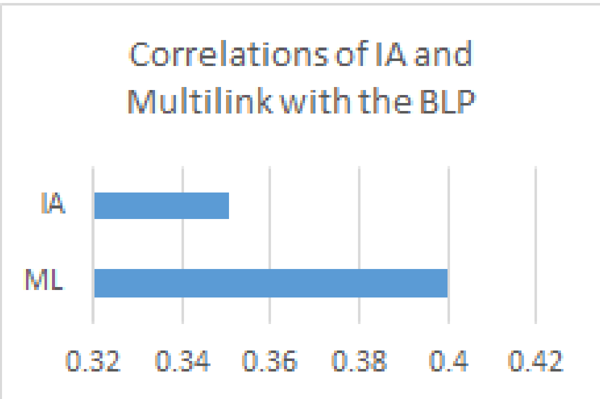

Figure 16 Correlations of the BIA and Multilink with the BLP

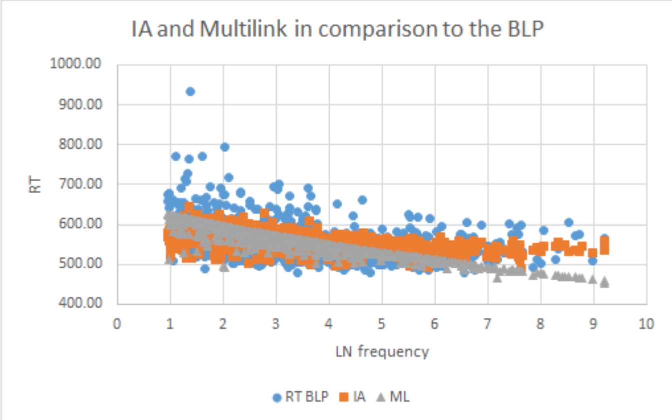

Figure 17 effect of frequency on the RTs of IA, Multilink, and BLP

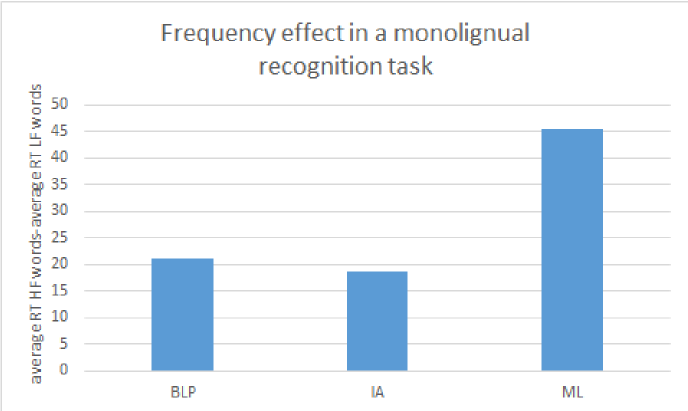

Figure 18 Frequency effects in Multilink, IA, and the BLP

8.3 Discussion

A first observation, is that Multilink 2016 underestimates the RTs for the very high frequencies. Where the curve of the BLP flattens for these frequencies, the Multilink RTs keep dropping. The effect of the high resting level activation of these words in the current Multilink parameterization seems to be too big, which makes that the high frequency words get translated too fast. This effect should thus be decreased in the high frequency words.

The second observation is that the frequency effect in Multilink is too large compared to the empirical data. However, correcting the underestimation of the RTs for very high frequency words will probably solve this problem as well. The average RTs for the high frequencies will become higher, and thus the difference between high frequency words and low frequency words will become smaller.

The third observation is the lack of variance in the Multilink data. On the one hand, the limited variability in the simulations does not match the empirical data, which makes the IA model performs better on this point. The IA model has an additional sub-lexical layer that is not implemented in Multilink (the letter layer) and this layer is probably responsible for the extra variance in the data of the IA model. Therefore, the implementation of an extra sub-lexical layer in Multilink should be considered, in order for Multilink to fit the empirical data better. On the other hand, the variance in the empirical data can be caused by individual differences in proficiency. Because we do not model individuals and their differences, the absence of the variance in Multilink would not a problem in this case.

9 Simulation 2: Word recognition in bilinguals

In the second simulation, I will investigate word recognition in bilinguals using the Multilink 2016 model and the BIA model (Dijkstra & Van Heuven, 1998). The simulation of bilinguals is the next step in the direction of word translation after word recognition in monolinguals.

9.1 Method

For this simulation, I used the online implemented BIA model by Van Heuven (http://www.psychology.nottingham.ac.uk/staff/wvh/index.html, 2015), and the same variant of the Multilink model that I discussed above. I adjusted the lexicon implemented in the BIA model to match the one used in Multilink, and I used all its Dutch non-cognate words to perform the recognition task (see appendix C). I compared the outcomes of both the BIA model and Multilink with the empirical data gathered by the Dutch Lexicon Project (DLP; Keuleers, Diependaele, & Brysbaert, 2010). In particular, I compared the frequency effect in the two models to that in the empirical data.

The high frequency and low frequency categories were obtained in the same way as in the first simulation. Similarly, the cycle times of Multilink and BIA were again adjusted to RTs using equation 8.

9.2 Results

Figure 19 shows the correlations between the DLP and the BIA, and the DLP and the Multilink model. Multilink performs considerably better than the BIA in terms of such correlations. Even though the shape of the frequency distribution in the BIA seems more in line with that in the empirical data, as can be seen in figure 20, Multilink reaches a correlation of nearly .60 with the DLP data, while the BIA shows a correlation of about .30.

Overall, we see the same pattern as in the simulation with the IA model (figure 17): again, when we plot the RTs against the frequencies, the slope in BIA seems to better fit that in the empirical data than that occurring in Multilink, and Multilink shows again less data variance than the data of the DLP and the output data of the BIA. However, Multilink shows a higher correlation to the empirical data than the BIA.

Finally, the frequency effect is again too large compared to the empirical data (figure 21), while the frequency effect of the BIA data seems to match that of the empirical data more closely.

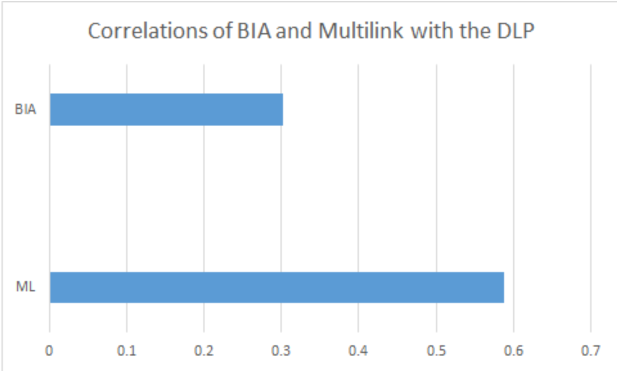

Figure 19 Correlations of BIA and Multilink with the DLP

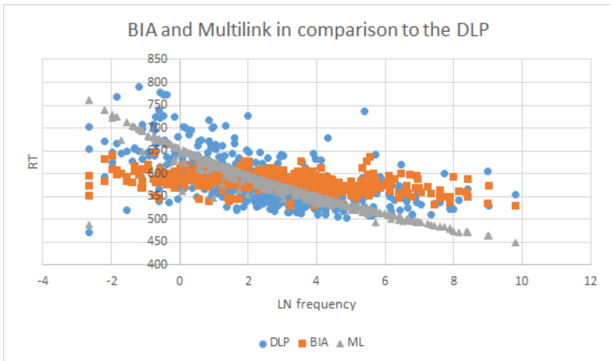

Figure 20 Effects of frequency on the reaction times on BIA, Multilink, and the DLP

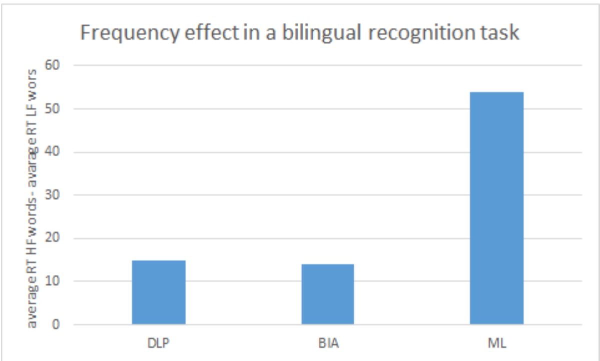

Figure 21 Frequency effect in DLP, BIA, and Multilink

9.3 Discussion

For bilingual word recognition, we can make the same observations in the output data of Multilink as in monolingual word recognition. On the one hand, the correlation of Multilink with the empirical data is larger than the correlation of the older BIA model to the empirical data. On the other hand, there is not enough variance in the Multilink data, there is an underestimation of RTs for high frequencies, and the overall frequency effect is too big.

10 Simulation 3: Word translation

The third simulation examines word translation. In this simulation, I will compare the adjusted Multilink model to the empirical data gathered by Pruijn (2015). In addition, I will focus on the frequency effect and the translation direction effect.

Earlier research with Multilink showed that the frequency effect in the model was too small (Peacock, 2015). The adjusted resting level activation range and the removal of the ranking system should increase the frequency effect. Besides the small frequency effect, no translation direction effect was found, where the empirical data by Pruijn (2015) showed a forward facilitation effect. To see if this finding arose because balanced rather than unbalanced bilinguals were simulated, I will simulate unbalanced bilinguals as well.

10.1 Method

For the last simulation, I used the translation task implemented in Multilink and compared the model outcomes to the empirical data by Pruijn (2015). For the simulations, I used the adjusted lexicon (see appendix A) and the same input words as Pruijn (2015) (see appendix D). Just as in Pruijn (2015), I distinguished 8 conditions by crossing the same 3 binary factors: cognate status (non-cognate or cognate), frequency (high frequency or low frequency), and translation direction (forward translation [Dutch to English] or backward translation [English to Dutch]).

First, I compared the RTs of the Multilink model (rescaled from cycle times via equation 8) with those of Pruijn (2015). Second, I focused on the frequency effects, again comparing that found by the model to that in the empirical data. Third, I looked at the translation direction effects of the Multilink data versus the empirical data.

In order to turn the English SUBTLEX (Brysbaert & New, 2009) frequencies into the subjective frequencies of unbalanced bilinguals, I applied three different transformations to the frequencies in the L2 lexicon: a linear transformation like Dijkstra et al. (1998) proposes, a non-linear transformation to 'punish' the low frequency words, and a non-linear transformation to 'punish' the high frequency words.

For the linear transformation, I divided the L2 frequencies by an arbitrary constant—I decided to use 4 and 10. For the non-linear transformation to 'punish' the low frequencies, I first squared the L2 frequencies. After this, I divided the values by 400 to rescale them (an arbitrarily chosen denominator). For the transformation to 'punish' the high frequency words, I transformed the values using the base 10 logarithm.

To see the influence of the unbalanced lexicons, I ran the same simulations with the different lexicons and compare their results.

10.2 Results

Figure 22 shows the average RTs for Multilink and the empirical data by Pruijn (2015), for the eight different conditions. The RT distribution in the Multilink output roughly follows the distribution found by Pruijn (2015). However, the frequency effect, and the translation direction effect both are too small in the Multilink outcomes compared to the empirical data, that can be seen in figure 33, that shows the frequency effects in Multilink and Pruijn (2015) and 36, that shows the . I will discuss this point in more detail in the next sections of this thesis.

The correlations between the Multilink outcomes and the data by Pruijn (2015) are shown in figure 23: forward translation shows a correlation of over .60 and backward translation a correlation of almost .50.

In figure 24 the correlations for forward translation and backward translation are split into a cognate category and a non-cognate category. The correlations for the non-cognates are better than the correlations for the cognates, which suggests that there is a flaw in the way cognates are determined or activated.

In figure 25 the correlations are split into the eight different categories Pruijn (2015) used, by adding a differentiation between low frequency words and high frequency words. Notably, is the negative correlation for forward translation in the low frequency cognates. Because the low frequent non-cognates do show a good correlation to the empirical data in forward translation, this implies that the problem is located in the implementation of the cognate facilitation effect, and not in the implementation of the frequency facilitation effect.

One observation is again the underestimated variance in the data points, especially in non-cognates. Figure 26 and 27 are boxplots of RTs from the empirical data, and show that the data obtained by Pruijn (2015) has a lot of variance, especially in the non-cognate category. Figure 28 and 29 are boxplots of RTs from the Multilink data, and show that this variance is absent in Multilink. This is also shown in more detail in figure 30 and 31, where the RTs of the non-cognates are plotted against the natural logarithm of the frequency. The Multilink data points all lay in a tight line, where the data points by Pruijn (2015) show a widespread cloud.

A problem occurring in Multilink is that still not all the words are translated correctly (see table 3), due to the cognate problem I described earlier. Instead of the right translation, Multilink finds a translation that has a lower Levenshtein distance to the input word than the correct translation has to the input word. Figure 32 shows what happens when translating the Dutch word AREND. Besides the right translation EAGLE (igP), the words FRIEND (frEnd) and GROUND (gr6nd) also get activated. The Levenshtein distance between AREND and FRIEND is just 2, which causes this word to be recognized more quickly.

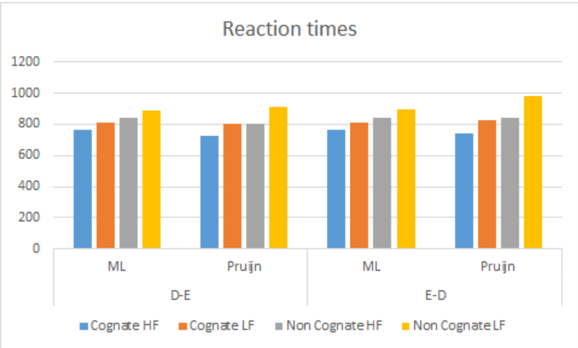

Figure 22 Reaction times for Pruijn (2015) and Multilink

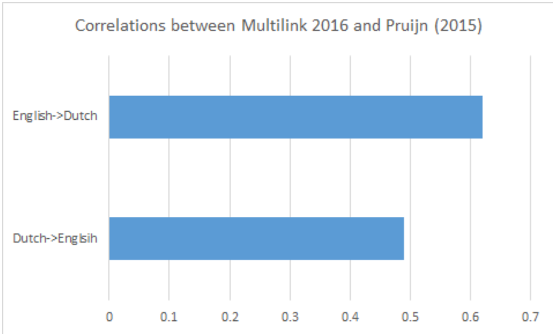

Figure 23 Correlation of Multilink cycle times and RTs by Pruijn (2015)

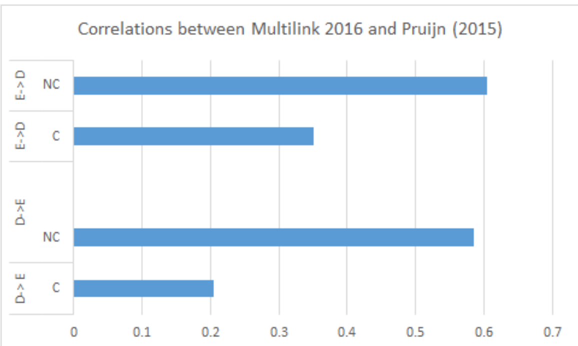

Figure 24 Correlation of Multilink cycle times and RTs by Pruijn (2015) for cognates and non-cognates

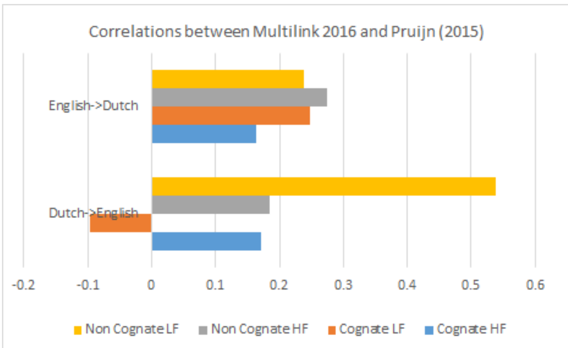

Figure 75 Correlation of Multilink cycle times and the RTs by Pruijn (2015) per category

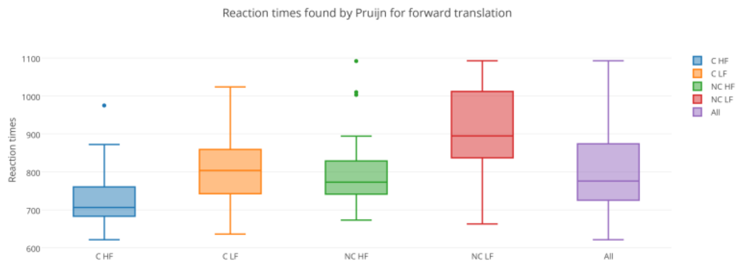

Figure 26 Boxplots of the reaction times for forward translation of the data by Pruijn (2015)

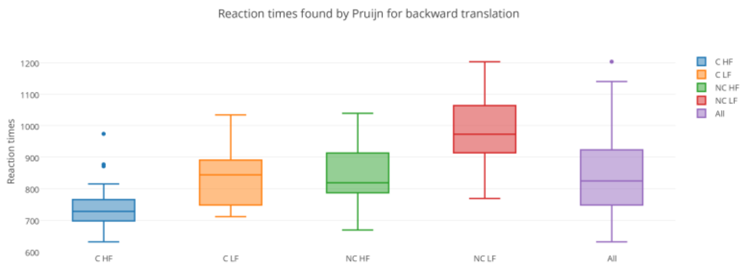

Figure 27 Boxplots of the reaction times for backward translation of the data by Pruijn (2015)

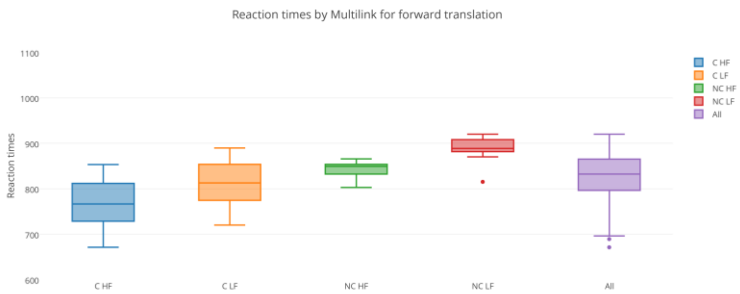

Figure 28 Boxplots of the reaction times for forward translation of the data by Multilink

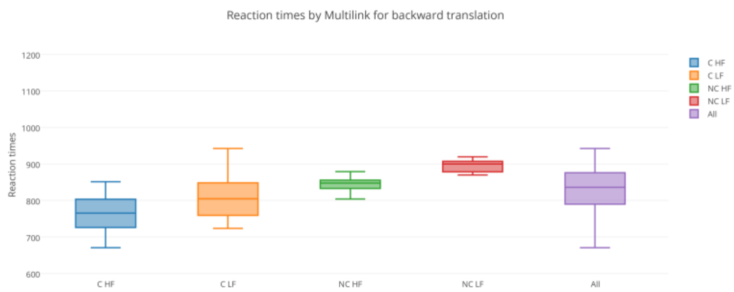

Figure 29 Boxplots of the reaction times for forward translation of the data by Multilink

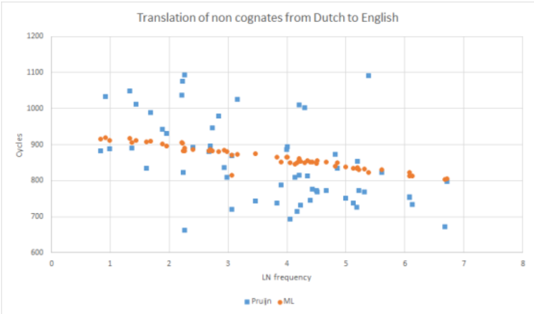

Figure 30 Spreading of the Multilink and Pruijn (2015) data of non-cognates in forward translation

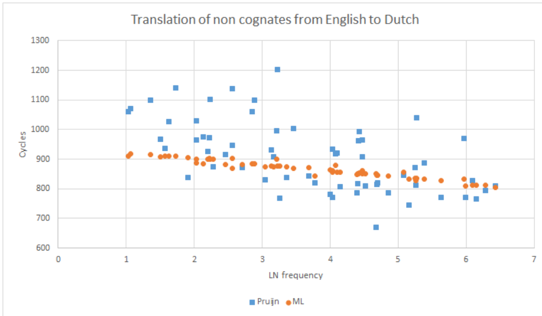

Figure 31 Spreading of the Multilink and Pruijn (2015) data of non-cognates in backward translation

Accuracies	D-E	E-D
C HF	100%	100%
C LF	97%	100%
NC HF	100%	100%
NC LF	81%	97%

Table 3 Accuracies for the Multilink model

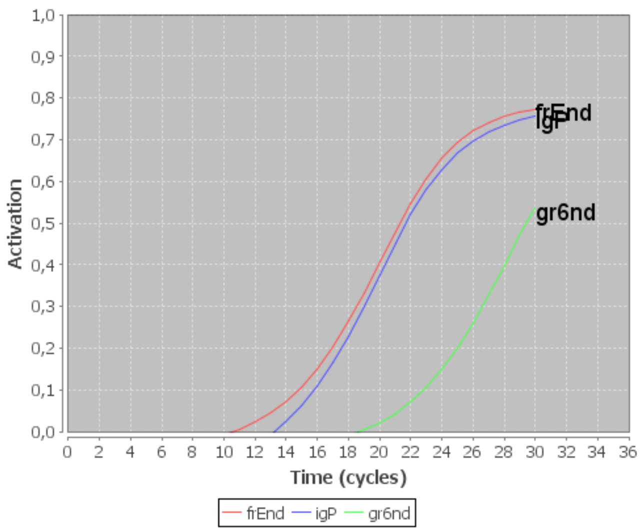

Figure 32 Activation for the input word AREND in a Multilink translation task

- Frequency effects

The data by Pruijn (2015) shows a frequency effect that is larger for non-cognates than for cognates, and is bigger in backward translation than in forward translation. This is shown in 33, where the frequency effects for the different categories are shown, for the Pruijn (2015) data and the Multilink data. The frequency effect in Multilink 2010 and Multilink 2016 presented in figure 34 show that the frequency effect has increased compared to the 2010 version of the model. However, it is still smaller than the effect in Pruijn (2015) (see figure 33). The differences between frequency effects in forward and backward translation, and between cognates and non-cognates, are also still too small in the Multilink data compared to the empirical data.

If we look at the cycle time differences in figure 34, we do see a small difference between cognates and non-cognates. Specifically, the frequency effect is bigger for the cognates in forward translation. However, this is opposite to the frequency effect found in the empirical data, where the effect is bigger for backward translation.

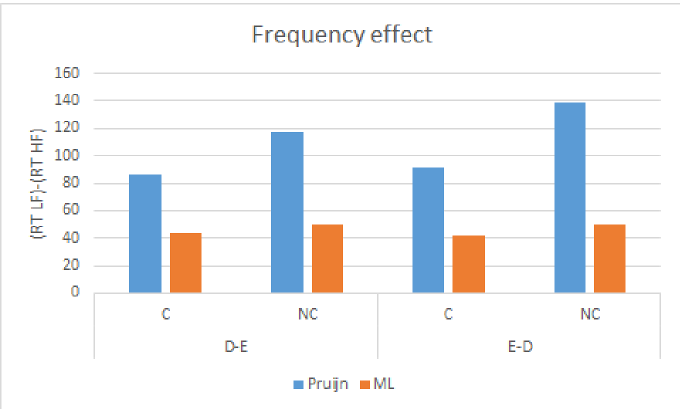

Figure 33 Frequency effects in Pruijn (2015) and Multilink

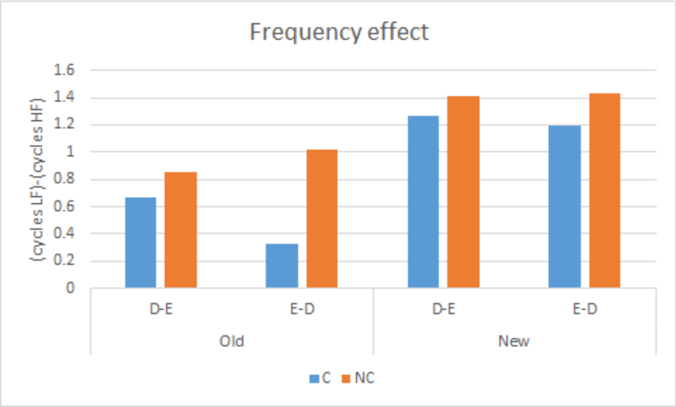

Figure 34 Frequency effect in the Multilink 2010 version and the 2016 Multilink model

- Translation direction effect

As mentioned earlier, Pruijn (2015) found a forward translation facilitation effect (see figure 35), but no translation direction effect was found in earlier research with Multilink. This is also what I found in the revised model and is shown in the most left graphs of figure 36. In this figure the difference between the cycle times for forward translation and backward translation are shown in the different categories and for the different transformations.

It is interesting that there is a small forward facilitation effect for the non-cognates, while the cognates show a small backward facilitation effect. It is not clear why this backwards facilitation effect shows up in the cognates, but it could occur because the average frequencies for the English cognates are somewhat higher than the frequencies of the Dutch cognates. This gives the English cognates a small advantage in recognition, which makes that the English words are translated faster. However, this does not explain the backward facilitation effect in the cognates with the lowered English frequencies. The reduction should have made the recognition of the English words (cognates and non-cognates) slower. This contradiction needs further research in order to be resolved.

The transformations to the lexicon by a division with 4 and 10, as proposed by Dijkstra et al. (1998), created a larger forward facilitation effect for the non-cognates. The effect was bigger when using the division by 10 than when using the division by 4, but the effect still remained too small compared to the data by Pruijn (2015) (figure 36). The difference between the cycles times stays under 0.4 cycles, which is not more than 10 milliseconds, adapted to latency times. The division mostly affected the high frequency words, where in the data by Pruijn (2015) the asymmetry is bigger for low frequency words. As expected, applying the logarithmic transformation increases the effect in the high frequency words, because the logarithmic transformation 'punishes' smaller numbers less than larger numbers. In contrast, taking the square and dividing the outcome by a constant 'punishes' the low frequency words more, which results in a bigger direction effect for the low frequency words.

The transformations did change the correlations to the empirical data, which are shown in figure 37 and figure 38. These graphs demonstrate that most of the correlations stay the same or go down when applying a transformation to the L2 lexicon. There are two exceptions: first, the correlation for high frequency cognates in the forward translation direction increases when applying the logarithmic transformation and square transformation. Second, the correlation for high frequency non-cognates in the backward translation direction increases after all transformations.

Figure 36 shows that none of the transformations change the backward facilitation effect occurring in the cognates. However, the backward facilitation effect does also increase for the cognates, when lowering the L2 frequencies in each of the transformations.

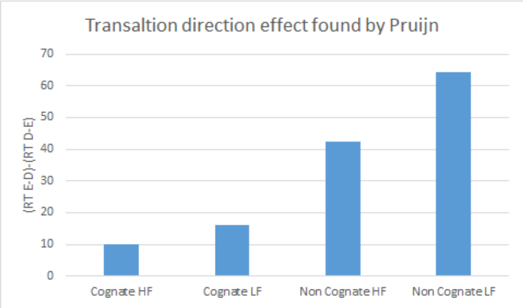

Figure 35 Translation direction effect found by Pruijn (2015)

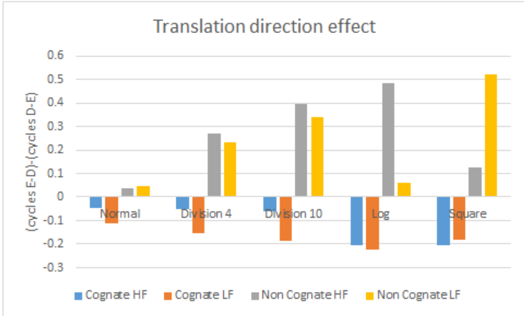

Figure 36 Translation direction effect found in Multilink for different L2 frequencies

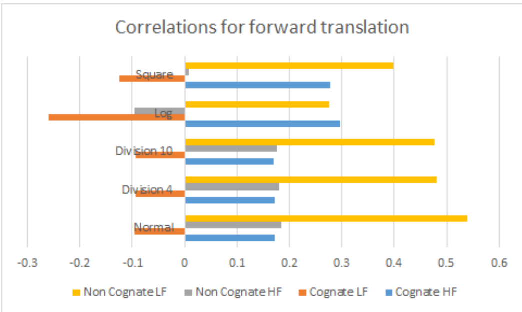

Figure 37 Correlations for Multilink output with data by Pruijn (2015) for lexicons with different L2 frequencies in forward translation

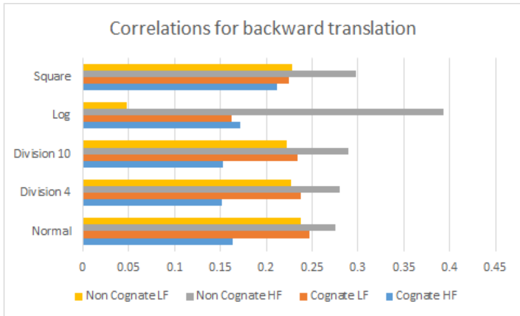

Figure 38 Correlations for Multilink output with data by Pruijn (2015) for lexicons with different L2 frequencies in backward translation

10.3 Discussion

The first series of word translation simulations show that there are still some problems in the Multilink 2016 model with respect to the effect of the cognate boost. Some of the words are still translated incorrectly, even though we have already reduced the effect of the cognates. De Korte (2016) found that the incorrect translations occur when the activation of the semantic node of a cognate pair with an incorrect translation overtakes the activation of the semantic node of the correct translation. Therefore, reducing the effect of the cognate status on the semantic node will solve this problem.

We saw that also in the translation task, Multilink shows less variance than the empirical data. However, as I mentioned before, the variance in the empirical data can also be caused by individual differences, which should not be modeled.

A third observation, is the negative correlation for the low frequency cognates in forward translation. Like I mentioned briefly in the results section, this negative correlation could be caused by just the cognate facilitation effect, because the non-cognates showed a better correlation. However, this could also show because the overestimations and underestimations of the RT in Multilink, caused by the cognate effect, frequency effect, and translation direction effect even each other out in other categories, but not in this one. For example, in the category forward translation of high frequent cognates, the too large cognate effect is evened out by the too small frequency effect. In the low frequency cognates category, the strong cognate effect is not evened out by a frequency effect, and therefore shows a low correlation. The high frequent non-cognates do not show a better correlation than the low frequent cognates, which would suggest that the overestimation of the cognate effect is stronger than the underestimation of the frequency effect. Because the low frequent non-cognates are not affected by a not optimal functioning facilitation effect, this category shows the best correlation to the empirical data.

In the second series of word translation simulations, the frequency effect did increase compared to the previous version of Multilink. However, it was still too small compared to the empirical data. The difference between the frequency effect in cognates and non-cognates, and the frequency effect in the different translation directions is also too small. This could be resolved by choosing a lower minimum value for the resting level activation, because this increases the effect of frequency. However, in the previous simulations we found that the frequency effect in the recognition task was too big, and lowering the minimum value will cause a further increase in the frequency effect in the recognition task as well. This needs more research before this can be applied.

In the last series of translation simulations, no translation direction asymmetry was found, in contrast to the forward facilitation effect by Pruijn (2015). The transformations to the L2 lexicon did have an effect on the translation direction effect: the effect became bigger, but still remained too small in any of the transformations. Most of the correlations between the Multilink data and the data from Pruijn (2015) decreased when the transformations were applied. Therefore, just the simple application of a transformation to the L2 lexicon is not enough to simulate a translation direction effect with Multilink. This could be because a transformation applied to the L2 lexicon is not an adequate way to simulate the subjective usage of the L2 by unbalanced bilinguals, or that the specific transformations I chose were not a correct way to simulate this usage. For example, the non-linear logarithmic function I used to simulate lower L2 proficiency caused that in total two logarithmic functions were applied on the frequency: one to simulate the L2 proficiency and one to calculate the resting level activations. Therefore, another transformation to simulate the lower L2 proficiency might be better.

However, it might also reflect a more fundamental problem and it could be that, in order to create an asymmetry in the output data, a different or additional representation is needed in the Multilink model, for example by building in a delay for the translation of L2 words.

However, because there is also no consistency across different empirical studies to the translation direction effect, the occurrence of a translation symmetry is not necessarily a problem. There is also empirical evidence for the translation symmetry currently showing in the Multilink model, for example, De Groot et al. (1994).

In all simulations, cognates performed less than non-cognates. First, the correlation with the data from Pruijn (2015) is low for the high frequency cognates, and the correlation for the low frequency cognates in forward translation direction is even negative. Second, the frequency effect in cognates in the forward translation direction is bigger than in backward translation, which is exactly the opposite from the empirical data. Third, the cognates show a backward translation facilitation effect instead of the forward facilitation effect found in the empirical data.

The translation of cognates is therefore an unsolved problem for the model. Solving this problem will probably also change the outcomes for the model regarding the translation direction effect for cognates and the frequency effect for cognates.

11 General discussion

The goal of my study was to investigate the frequency effect and the translation direction effect in the Multilink 2016 model. Furthermore, I wished to resolve the issues around these effects found in previous done research. To accomplish these aims, I conducted three simulation studies with a revised version of the Multilink model: a monolingual word recognition simulation, a bilingual word recognition simulation, and a word translation simulation.

The first two simulations concerned word recognition. First, I simulated monolingual word recognition with the IA and the 2016 version of Multilink, and compared their outcomes to the empirical data of the BLP. Second, I performed a bilingual recognition simulation using the BIA and the 2016 version of Multilink, and I compared their outcomes to the empirical data by the DLP.

With these simulations Multilink already goes beyond the RHM. In contrast to the RHM, which outcomes are just based on intuitions, we can compare the concrete Multilink outcomes to empirical data, and study the similarities and differences.

The results of the two word recognition simulations showed the same patterns. The correlation between Multilink and the empirical data was higher than the correlation between the older models (IA and BIA) and the empirical data, in both simulations. The frequency effect in the Multilink data overestimated that in the empirical data. Furthermore, the Multilink data did not contain enough variability compared to the data of the BLP and DLP.

The third simulation concerned word translation, for which I used the 2016 version of Multilink, and compared the outcomes to the empirical data on word translation gathered by Pruijn (2015). In this simulation Multilink not only goes beyond the RHM, but also passes the IA and the BIA, because the these models are only capable of simulating recognition.

The results of the comparison between Multilink and the empirical data showed some good correlations, especially for the non-cognates. Thereby, the frequency effect in the 2016 version of Multilink has increased compared to the 2010 version by lowering the minimum resting level activation, but it was too small compared to the frequency effect found by Pruijn (2015). Lastly, a translation symmetry occurred when simulating unbalanced bilinguals, as well as when simulating balanced bilinguals, where the data by Pruijn (2015) showed a forward facilitation effect.

In the simulations I found that the frequency effect is too big in the recognition tasks and too small in the translation task. This suggests that the range of the resting level activation, which is currently at -.20 to 0, probably needs to be adjusted to decrease or increase the frequency effect. The large frequency effect in the recognition simulations was due to the underestimation of the RTs for the high frequency words. Because earlier research (Peacock, 2015) showed that these words caused unwanted complications, we left out the very high frequent words in the translation task. As a consequence, because the translation simulations did not have the extremely low reaction times for high frequency words, this underestimation did not appear in the translation task. This also caused the smaller frequency effect in the translation task than in the recognition task.

Therefore, it is important that high frequency words are simulated in the translation tasks as well, before changing the range of the resting level activation, in order to reach an unequivocally conclusion about the high frequency words. If the translation task shows the same underestimation for these words, this needs to be solved first, by changing the way the resting level activation is calculated. Because applying the log frequency as a way to calculate the resting level activation caused a decrease in some of the correlations with the empirical data, further research to using another range, that ensures that the high frequency words are translated somewhat slower than when using the logarithm, would be a good place to start.

A translation direction effect did not arise in the Multilink data, and the effect of the transformations (division, logarithm, square) to the L2 lexicon, in order to simulate unbalanced bilinguals, were not large enough compared to the data by Pruijn (2015). Apparently, the effect of the faster recognition of Dutch words is evened out by the slower retrieval of the corresponding English word. In order to overcome this, the parameters could be changed, for example by making a distinction between the two routes. When the connections on the forward translation route (from Dutch orthography, to semantics, and to English phonology) are somewhat stronger than the connections on the backward translation route (from English orthography, to semantics, and to Dutch phonology), forward translation is advantaged, which will cause a forward translation facilitation effect.

However, solving the problem of the insufficient frequency effect in the translation task could be enough to create a translation direction effect. Namely, dividing the L2 frequencies by a constant makes almost all L2 words low frequency. If the frequency effect becomes larger, low frequency words are translated slower and high frequency words are translated faster. Because the consequences of lowering the frequencies are relatively large, the translation direction effect might also increase. On the contrary, because the available empirical data about the translation direction effect is not consistent, it is not necessarily a problem to the model that a translation symmetry is produced.

Regarding the frequency effect and the translation direction effect, the cognates represent a problem on their own. They do show the cognate facilitation effect, but they do not show other patterns occurring in the empirical data, such as the forward facilitation effect. Because the absence of these effects do not show in the non-cognates, it is probably dependent on the implementation of the cognate facilitation effect. This indicates that, even though the cognate facilitating effect has already been reduced, it is still too strong, and therefore overpowers the other effects, like the frequency effect. In order to solve this problem, experimenting with an even more reduced cognate effect will be useful.

In none of the simulations, the Multilink data showed as much variance as the empirical data. This lack of variance is not necessarily a problem, as long as it does not have a systematic source that should be accounted for in the model. Nevertheless, it would be good to see if implementing attributes that increase the data variance have a positive effect on the performance of the model, such as a latency cost for different phoneme onset times or an extra sub-lexical layer.

The BIA and the IA models showed somewhat more data variance, which is probably due to the extra sub-lexical layer (the letter level) in these models. However, the absence of this layer in Multilink gives it the advantage of processing words of different lengths, where the IA and BIA model are limited to 4-letter words. It would be ideal to find a way to increase systematic data variance without limiting the model to a strict length of the input words. Experimenting with sub-lexical layers would be one way to see if this indeed changes the variability of the output data.

In sum, several studies (e.g. Palo, Schaeffler, and Scobbie, 2015) have found that phoneme onsets influence the RTs. Because some phonemes are harder to produce articulatory, these phonemes take longer to pronounce than others. Including this effect in Multilink would probably have a positive effect on the degree of variance in Multilink output, because more individual aspects of words are taken into account in translation of a word. As a consequence, the RTs of the words will become more distinct, and the overall RT distribution will, therefore, show more variance.

11.1 Future research

Future research needs to be done to resolve the above-mentioned issues. First, research to create more systematic variance in the Multilink model is needed. Important here is to see if this really improves the model, or that the data variance in the empirical data is due to individual proficiency differences. In the latter case, we do not want and need to simulate the variance in the empirical data. To create more variance in the Multilink output data, one option is to add an extra sub-lexical layer, like in the IA and

BIA model. Another option to create more variability in the phonological Multilink output data, is to add in a latency cost for onset phonemes that are hard to produce.

Second, because the cognates performed worse than the non-cognates in the model, there needs to be extra research done regarding the cognates. This research should concern the use of the Levenshtein distance as similarity measure and the score function that gives more similar words a boost, to activate them more quickly.

Third, future studies should consider the translation direction effect in more detail. Currently, no asymmetry is observed in the Multilink data, in contrast to the empirical data by Pruijn (2015). Unfortunately, the effect of changing the L2 word frequencies was not big enough in this regard. However, because there is still so much argument about the existence of the translation direction effect, this first needs to be resolved before we can draw any conclusion about the occurrence of the translation direction effect in Multilink.

Fourth, the simulated frequency effect in word recognition was too large, because the cycle times from the Multilink output are too low for high frequency words in the recognition task. Because the range of the resting level activation, and the calculation of the resting level activations are the base of the frequency effect in Multilink, these two aspects need to be adjusted in order to solve this underestimation, for example by choosing another distribution than the logarithm to calculate the resting level activations. These high frequency words were not modeled in the translation task, therefore it is not sure if the translation task also shows this underestimation of the high frequency words. This is also important to study in further research.

Fifth, future studies should focus on the calculation of the resting activations of the words in the lexicon. We changed the old ranking system to one where base 10 logarithm of the frequencies is used to determine the resting level activation. Even though this method takes in account the scalar differences between the frequencies of 2 words in the lexicon, some of the correlations were reduced. It could be that using another scale than the logarithmic scale provides better correlations to the empirical data.

Finally, De Korte (2016) found that the model-to-data fit to the data by Pruijn (2015) increased when reducing the minimum resting level activation to -.40, which increases the frequency effect in Multilink. However, because Multilink already showed an underestimation of the RTs for the high frequency words in the recognition simulations with the minimum value -.20, and because the decrease could worsen this underestimation, further research should show if this effect will not be compounded by reducing the minimum resting level activation.

12 Conclusion

In previous studies with the Multilink 2010 model, such as those by Lormans (2012) and Peacock (2015), the 2010 version of the model showed potential in the field of word translation, but problems with respect to three important translation effects: the frequency effect, the translation direction effect, and the cognate effect.

The goal of this thesis was to improve to model of the basis of the following two research questions:

1. What modifications to the resting level activation can be done to increase the frequency effect in the model?
2. Can the simulation of unbalanced bilinguals increase the forward translation effect in the Multilink model?

To improve the performance of the model, it was adapted to a new version. This revised version of the model includes phonological representations, an adjusted lexicon, uses a different score measure to simulate cognate effects, and applies the log frequency instead of a ranking system to determine the resting level activations of words.

I conducted three simulation studies to answer the research questions by studying the frequency effect and the translation direction effect in the 2016 version of the Multilink model: a monolingual recognition task, a bilingual recognition task, and a translation task.

In both monolingual and bilingual model-to-model simulations, Multilink produced higher correlations to the empirical data than the earlier IA model and BIA model. On the down side, the frequency effect in these simulations was too large, and the variance in the simulation data was much smaller than that in the empirical data.

The translation simulation showed that the frequency effect in the output of the 2016 version of model was increased compared to the 2010 version, but still too small compared to the empirical data. Furthermore, translation shows a symmetrical effect between languages, instead of a forward facilitation effect (faster translation from L1 to L2 than vice versa) as found by Pruijn (2015).

These findings provide the following answers to my research questions: the decreased minimum value of the resting level activation has increased the frequency effect in the model, however compared to the empirical data it is still too small. De Korte (2016) found that reducing the minimum resting level activation to -.40 provided a better model-to-data fit to the empirical translation data by Pruijn (2015). It will be interesting to see how this effects the high frequency words in word recognition that already showed a too low RT with -.20 as minimum resting level activation.

Thereby, we changed the ranking system to a system using the logarithm of the frequency, to take the real frequency differences between words in account. However, since some of the correlations dropped and RTs for the high frequency words in recognition were too low, further research could experiment with other distributions.

Simulating unbalanced bilinguals instead of balanced bilinguals is not enough to create a translation asymmetry. Further research could focus on an asymmetric parameterizing in order to create an asymmetry in translation. However, because there is no common agreement on the translation direction effect in the experimental data, more empirical research to this effect is needed to provide an unambiguously conclusion.

Altogether, is Multilink the first computational translation model that takes in account an orthographic, semantic, and phonological component, which makes it an revolutionary model in the field of psycholinguistics. Thereby, shows Multilink promising results in becoming an all-round model for word recognition and word translation in balanced and unbalanced bilinguals, because Multilink is not only capable of just simulating a translation task, it also shows important linguistic effects such as the frequency effect and the cognate effect. However, in order to capture essential aspects of these processes more precisely extra research is required with respect to the translation direction effect, the frequency effect, and the cognate facilitation effect. In sum, the earlier presented results of the simulation with the 2016 version of Multilink show the added value of the model to the psycholinguistic field.

13 References

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition*, *13*(03), 359-371.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior research methods*, *41*(4), 977-990.

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. Behavior Research Methods, 44(4), 991-997.

Burnage, G., Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1990). *CELEX: A guide for users*. CELEX.

Christoffels, I. K., De Groot, A. M., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, *54*(3), 324-345.

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. Psychonomic bulletin & review, 22(5), 1216-1234.

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al (1991).

Degroot, A., Dannenburg, L., & Vanhell, J. (1994). Forward and Backward Word Translation by Bilinguals. Journal Of Memory And Language, 33(5), 600-629.

Dijkstra, T., & Rekké, S. (2010). Towards a localist-connectionist model of word translation. ML, 5(3), 401-420.

Dijkstra, T., & van Heuven, W. (2002). The architecture of the bilingual word recognition system: From identification to decision. Bilingualism, 5(03).

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. Journal of Memory and Language, 62, 284-301.

Dijkstra, T., & Van Heuven, W. J. (1998). The BIA model and bilingual word recognition. Localist connectionist approaches to human cognition, 189-225.

Dijkstra, T., Lormans, E., Rekké, S., & Roelofs, A. Multilink: a computational model for word translation

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review*, *15*(4), 850-855.

Duyck, W., & Warlop, N. (2009). Translation priming between the native language and a second language: New evidence from Dutch-French bilinguals. Experimental Psychology, 56(3), 173-179.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. Behavior Research Methods, 42(3), 643-650.

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono-and disyllabic words and nonwords. *Frontiers in Psychology*, *1*(18 november), 1-15.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287-304.

Kroll, J., & Stewart, E. (1994). Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections Between Bilingual Memory Representations. Journal Of Memory And Language, 33(2), 149-174.

Kroll, J., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. Second Lang Res, 18(2), 137-171.

Levenshtein, V. I. (1966, February). Binary codes capable of correcting deletions, insertions, and reversals. In Soviet physics doklady (Vol. 10, No. 8, pp. 707-710).

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. Cognition, 25(1-2), 71-102.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings.*Psychological review*, *88*(5), 375.

McLuhan, M. (1962). The Guttenberg Galaxy. *The making of topografic man*. Toronto, Canada: University of Toronto Press.

Morton, J. (1970). A functional model of human memory. In D. A. Norman (Ed.), Models of human memory (pp. 203-254). New York: Academic Press

Palo, P., Schaeffler, S., & Scobbie, J. (2015). Effect of phonetic onset on acoustic and articulatory speech reaction times studied with tonue ultrasound.

Peacock J. (2016). Modelling Lexical Effects with Multilink: Frequency, Cognate Status, and Translation Asymmetry (Master Thesis). Radboud University Nijmegen.

Pruijn, L. (2015). From ‘Daar’ to ‘There’ and Back Again: About the Mechanisms underlying Word Translation (Master Thesis). Radboud University Nijmegen.

Rekke, S.T. (2010). Multilink: a model for bilingual processing (bachelor thesis). Radboud University Nijmegen

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological review*, *89*(1), 60.

Taft, M., & Hambly, G. (1986). Exploring the cohort model of spoken word recognition. *Cognition*, *22*(3), 259-282.

VAN HELL, J. G., TOKOWICZ, N., & GREEN, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment.

Van Heuven, W. J., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of memory and language*, *39*(3), 458-483.

Wit, de W. (2014). Exploration and Testing of the Multilink Model Simulating the Word Translation Process (Bachelor Thesis) Radboud University Nijmegen

14 Appendix

14.1 Appendix A: Multilink lexicon

Dutch:O	Dutch:P	English:O	English:P
AANDEEL,9.95	andel,9.95	SHARE,69.51	S8R,69.51
AANLEG,2.88	anlEx,2.88	INSTANCE,16.78	Inst@ns,16.78
AARDE,100.07	ard@,100.07	EARTH,99.49	3T,99.49
AARDIG,191.95	ard@x,191.95	FRIENDLY,26.04	frEndlI,26.04
AAS,9.06	as,9.06	BAIT,9.73	b1t,9.73
ACHTER,473.8	Axt@r,473.8	BEHIND,187.86	bIh2nd,187.86
ACHTING,1.07	AxtIN,1.07	ESTEEM,1.69	Istim,1.69
ACTEUR,20.99	Akt|r,20.99	ACTOR,26.33	{kt@R,26.33
ACTIE,39.81	Aksi,39.81	ACTION,61.08	{kSH,61.08
ADEM,53.15	ad@m,53.15	BREATH,44.92	brET,44.92
ADVIES,33.23	Atfis,33.23	ADVICE,47.88	@dv2s,47.88
AFGUNST,1.56	Afx}nst,1.56	ENVY,9.55	EnvI,9.55
AFSCHEID,35.45	AfsxKt,35.45	GOODBYE,116.51	gUdb2,116.51
AFSCHUW,1.62	Afsxyw,1.62	HORROR,9.18	hQr@R,9.18
AFSTAND,39.97	AfstAnt,39.97	DISTANCE,25.61	dIst@ns,25.61
AFVOER,2.4	Afur,2.4	DRAIN,8.63	dr1n,8.63
AFWAS,3.82	AfwAs,3.82	DISHES,11.86	dISIz,11.86
ALARM,34.78	alArm,34.78	ALARM,29.84	@l#m,29.84
ALLEEN,1469.24	Alen,1469.24	ALONE,308.53	@l5n,308.53
ALTIJD,978.45	AltKt,978.45	ALWAYS,655.25	$lw1z,655.25
ANDER,190.22	And@r,190.22	OTHER,735.39	VD@R,735.39
ANGEL,10.63	AN@l,10.63	STING,7.02	stIN,7.02
ANGST,69.34	Anst,69.34	FEAR,69.08	f7R,69.08
ANKER,5.03	ANk@r,5.03	ANCHOR,7.41	{Nk@R,7.41
ANTWOORD,140.11	Antwort,140.11	ANSWER,176.2	#ns@R,176.2
AREND,2.06	ar@nt,2.06	EAGLE,11.49	igP,11.49
ARM,79.26	Arm,79.26	POOR,129.08	p$R,129.08
ARMBAND,6.27	ArmbAnt,6.27	BRACELET,7.8	br1slIt,7.8
ATLEET,2.58	Atlet,2.58	ATHLETE,4.61	{Tlit,4.61
ATOOM,1.14	atom,1.14	ATOM,2.75	{t@m,2.75
AUTEUR,4.12	Mt|r,4.12	AUTHOR,7.94	$T@R,7.94
AUTO,458	Mto,458	CAR,483.06	k#R,483.06
AVOND,180.61	avOnt,180.61	EVENING,120.69	ivHIN,120.69
BAAI,4.51	baj,4.51	BAY,24.24	b1,24.24
BAAL,3.96	bal,3.96	SACK,12.92	s{k,12.92

BAAN,158.45	ban,158.45	JOB,413	_Qb,413
BAARD,11.64	bart,11.64	BEARD,12.61	b7d,12.61
BAAS,167.21	bas,167.21	BOSS,124.29	bQs,124.29
BAAT,2.47	bat,2.47	BENEFIT,14.35	bEnIfIt,14.35
BABY,151.8	bebi,151.8	BABY,509.37	b1bI,509.37
BAKKER,4	bAk@r,4	BAKER,13.69	b1k@R,13.69
BAL,80.63	bAl,80.63	BALL,104.96	b$l,104.96
BALKON,6.63	bAlkOn,6.63	BALCONY,7.31	b{lk@nI,7.31
BALLON,5.28	bAlOn,5.28	BALLOON,8.67	b@lun,8.67
BANAAN,5.33	banan,5.33	BANANA,10.73	b@n#n@,10.73
BAND,79.28	bAnt,79.28	TIRE,12.37	t2@R,12.37
BANK,91.91	bANk,91.91	COUCH,23.47	k6J,23.47
BASIS,34.71	baz@s,34.71	BASE,35.37	b1s,35.37
BEDRAG,14.73	b@drAx,14.73	AMOUNT,24.75	@m6nt,24.75
BEDRIJF,73.38	b@drKf,73.38	COMPANY,147.2	kVmp@nI,147.2
BEEK,3.43	bek,3.43	BROOK,2.04	brUk,2.04
BEELD,40.06	belt,40.06	STATUE,10.59	st{Ju,10.59
BEHAARD,0.73	b@hart,0.73	HAIRY,6.31	h8rI,6.31
BEKEND,89.8	b@kEnt,89.8	KNOWN,123.53	n5n,123.53
BEKER,8.78	bek@r,8.78	CUP,51.65	kVp,51.65
BEL,311.67	bEl,311.67	BUBBLE,8	bVbP,8
BELANG,32.27	b@lAN,32.27	INTEREST,50.94	Intr@st,50.94
BELEEFD,17.49	b@left,17.49	POLITE,13.94	p@l2t,13.94
BELEID,9.99	b@lKt,9.99	POLICY,27.02	pQl@sI,27.02
BERICHT,59.78	b@rIxt,59.78	MESSAGE,91.51	mEsI_,91.51
BEROEMD,24.49	b@rumt,24.49	FAMOUS,45.02	f1m@s,45.02
BES,0.62	bEs,0.62	BERRY,3.49	bErI,3.49
BESLUIT,28.26	b@slLt,28.26	DECISION,55.06	dIsIZH,55.06
BETON,4.34	b@tOn,4.34	CONCRETE,7.43	kQNkrit,7.43
BEWIJS,101.92	b@wKs,101.92	PROOF,34.39	pruf,34.39
BIEFSTUK,7.36	bifst}k,7.36	STEAK,16.24	st1k,16.24
BIER,53.35	bir,53.35	BEER,75.49	b7R,75.49
BIJBEL,23.05	bKb@l,23.05	BIBLE,18.33	b2bP,18.33
BIJL,9.26	bKl,9.26	AXE,4.88	{ks,4.88
BIJSTAAN,4.07	bKstan,4.07	ASSIST,7.84	@sIst,7.84
BIJVAL,0.25	bKvAl,0.25	APPROVAL,8.39	@pruvP,8.39
BISSCHOP,9.88	bIsxOp,9.88	BISHOP,16.76	bIS@p,16.76
BLAD,11.14	blAt,11.14	LEAF,5.2	lif,5.2
BLEEK,24.17	blek,24.17	PALE,8.02	p1l,8.02

BLIJVEN,517.57	blKv@,517.57	REMAIN,33.22	rIm1n,33.22
BLOED,185.32	blut,185.32	BLOOD,186.12	blVd,186.12
BLOEM,13.49	blum,13.49	FLOWER,22.76	fl6@R,22.76
BLOESEM,0.66	blus@m,0.66	BLOSSOM,3.61	blQs@m,3.61
BOCHT,8.37	bOxt,8.37	BEND,15.06	bEnd,15.06
BODEM,15.3	bod@m,15.3	SOIL,7.78	s4l,7.78
BOEK,150.93	buk,150.93	BOOK,176.98	bUk,176.98
BOEKEN,66.91	buk@,66.91	ACHIEVE,7.33	@Jiv,7.33
BOEKET,2.06	bukEt,2.06	BOUQUET,3.22	bUk1,3.22
BOEKJE,11.02	bukj@,11.02	NOTEBOOK,4.61	n5tbUk,4.61
BOER,14.77	bur,14.77	FARMER,11.84	f#m@R,11.84
BOK,2.17	bOk,2.17	BUCK,33.75	bVk,33.75
BOKSER,6.45	bOks@r,6.45	BOXER,3.84	bQks@R,3.84
BOON,1.42	bon,1.42	BEAN,6.84	bin,6.84
BOOR,3.45	bor,3.45	DRILL,13.75	drIl,13.75
BOOS,105.79	bos,105.79	ANGRY,58.98	{NgrI,58.98
BOOT,95.93	bot,95.93	BOAT,95.78	b5t,95.78
BORD,27.3	bOrt,27.3	PLATE,25.65	pl1t,25.65
BORG,11.23	bOrx,11.23	SECURITY,94.31	sIkj9r@tI,94.31
BORSTEL,2.1	bOrst@l,2.1	BRUSH,14.16	brVS,14.16
BOT,14.52	bOt,14.52	BONE,26.06	b5n,26.06
BOTER,10.52	bot@r,10.52	BUTTER,20.43	bVt@R,20.43
BOUILLON,1.28	buljOn,1.28	BROTH,0.88	brQT,0.88
BOUT,0.94	bMt,0.94	BOLT,6.88	b5lt,6.88
BROEK,67.28	bruk,67.28	TROUSERS,5.16	tr6z@z,5.16
BROOD,33.59	brot,33.59	LOAF,4.47	l5f,4.47
BROODJE,12.76	brotj@,12.76	BUN,2.88	bVn,2.88
BRUID,21.13	brLt,21.13	BRIDE,24.22	br2d,24.22
BRUIN,13.68	brLn,13.68	BROWN,60.12	br6n,60.12
BUFFEL,1.12	b}f@l,1.12	BUFFALO,11.9	bVf@l5,11.9
BUIDEL,0.73	bLd@l,0.73	POUCH,1.71	p6J,1.71
BUIS,6.08	bLs,6.08	TUBE,16.43	tjub,16.43
BULT,3.41	b}lt,3.41	BUMP,12.35	bVmp,12.35
BUREAU,66.93	byro,66.93	DESK,43.9	dEsk,43.9
BURGER,19.19	b}rG@r,19.19	CITIZEN,13.33	sItIzH,13.33
BUS,64.83	b}s,64.83	MAILBOX,4.16	m1lbQks,4.16
CADEAU,29.29	kado,29.29	PRESENT,89.45	prIzEnt,89.45
CEL,49.97	sEl,49.97	CELL,54.35	sEl,54.35
CENTRAAL,2.68	sEntral,2.68	CENTRAL,24.75	sEntr@l,24.75

CHEMISCH,3.09	xemis,3.09	CHEMICAL,11.35	kEmIkP,11.35
CHIRURG,8.85	Sir}rx,8.85	SURGEON,16.43	s3_@n,16.43
CHLOOR,0.98	xlor,0.98	CHLORINE,0.75	kl$rin,0.75
CIRKEL,12.44	sIrk@l,12.44	CIRCLE,21.51	s3kP,21.51
COCON,1.42	kokOn,1.42	COCOON,1.12	k@kun,1.12
CONGRES,12.28	kONGrEs,12.28	CONGRESS,8.22	kQNgrEs,8.22
CRU,0.57	kry,0.57	CRUDE,3.04	krud,3.04
CULTUUR,11.62	k}ltyr,11.62	CULTURE,13.94	kVlJ@R,13.94
CURSUS,8.78	k}rz}s,8.78	COURSE,487.22	k$s,487.22
DAGBOEK,16.99	dAGbuk,16.99	DIARY,8.98	d2@rI,8.98
DAK,54.84	dAk,54.84	ROOF,35.65	ruf,35.65
DALING,1.19	dalIN,1.19	DECLINE,2.98	dIkl2n,2.98
DAME,82.74	dam@,82.74	LADY,217.08	l1dI,217.08
DAMP,0.41	dAmp,0.41	FUMES,2.86	fjumz,2.86
DANS,37.82	dAns,37.82	DANCE,148.04	d#ns,148.04
DAS,13.38	dAs,13.38	TIE,44.43	t2,44.43
DATUM,13.84	dat}m,13.84	DATE,141.53	d1t,141.53
DECOR,1.62	dekOr,1.62	SCENERY,3.78	sin@rI,3.78
DEEG,1.85	dex,1.85	DOUGH,15.88	d5,15.88
DEEL,142.35	del,142.35	PART,261.51	p#t,261.51
DERDE,76.49	dErd@,76.49	THIRD,74.53	T3d,74.53
DEUGD,6.43	d|xt,6.43	VIRTUE,5.16	v3tju,5.16
DEUK,2.65	d|k,2.65	DENT,3.53	dEnt,3.53
DEUR,247.48	d|r,247.48	DOOR,292.06	d$R,292.06
DICHT,182.3	dIxt,182.3	SHUT,263.82	SVt,263.82
DICHTER,27.24	dIxt@r,27.24	POET,9.22	p5It,9.22
DIEET,9.24	dijet,9.24	DIET,15.37	d2@t,15.37
DIEF,29.96	dif,29.96	THIEF,24.27	Tif,24.27
DIEFSTAL,13.51	difstAl,13.51	THEFT,6.75	TEft,6.75
DIENBLAD,1.23	dimblAt,1.23	TRAY,8.04	tr1,8.04
DIEP,68.4	dip,68.4	DEEP,76.39	dip,76.39
DIER,28.1	dir,28.1	ANIMAL,45.49	{nImP,45.49
DIK,42.6	dIk,42.6	THICK,13.98	TIk,13.98
DILLE,0.25	dIl@,0.25	DILL,1.76	dIl,1.76
DINER,24.08	dine,24.08	DINNER,202.67	dIn@R,202.67
DINGEN,510.78	dIN@,510.78	THINGS,692.88	TINz,692.88
DOCHTER,219.3	dOxt@r,219.3	DAUGHTER,171.35	d$t@R,171.35
DODEN,217.98	dod@,217.98	KILL,452.57	kIl,452.57
DOEK,8.83	duk,8.83	CLOTH,6.1	klQT,6.1

DOEL,77.45	dul,77.45	TARGET,37.96	t#gIt,37.96
DOK,7.66	dOk,7.66	DOCK,9.75	dQk,9.75
DOM,77.82	dOm,77.82	SILLY,57.1	sIlI,57.1
DOMEIN,3.11	domKn,3.11	DOMAIN,2.59	d5m1n,2.59
DONKER,64.44	dONk@r,64.44	DARK,88.61	d#k,88.61
DOOD,992.15	dot,992.15	DEATH,216.69	dET,216.69
DOOF,18.59	dof,18.59	DEAF,14.53	dEf,14.53
DOOLHOF,2.63	dolhOf,2.63	MAZE,2.55	m1z,2.55
DORP,53.99	dOrp,53.99	VILLAGE,33.57	vIlI_,33.57
DORSTIG,2.42	dOrst@x,2.42	THIRSTY,12.29	T3stI,12.29
DOUCHE,22.25	duS,22.25	SHOWER,41.12	S6@R,41.12
DOZIJN,7.91	dozKn,7.91	DOZEN,24.14	dVzH,24.14
DRACHT,0.14	drAxt,0.14	COSTUME,14.14	kQstjUm,14.14
DREIGING,5.81	drKGIN,5.81	THREAT,20.76	TrEt,20.76
DRINGEND,23.19	drIN@nt,23.19	URGENT,14.86	3_@nt,14.86
DROOG,19.64	drox,19.64	DRY,42.82	dr2,42.82
DRUK,207.11	dr}k,207.11	PRESSURE,53.12	prES@R,53.12
DUBBEL,17.08	d}b@l,17.08	DOUBLE,62.71	dVbP,62.71
DUIF,5.35	dLf,5.35	PIGEON,5.9	pI_In,5.9
DUIM,10.79	dLm,10.79	THUMB,11.82	TVm,11.82
DUIN,0.39	dLn,0.39	DUNE,1	djun,1
DUIVEL,45.12	dLv@l,45.12	DEVIL,41.33	dEvP,41.33
DUN,9.1	d}n,9.1	THIN,20.18	TIn,20.18
DWAAS,22.75	dwas,22.75	FOOLISH,17.51	fulIS,17.51
DWALING,0.62	dwalIN,0.62	ERROR,9.27	Er@R,9.27
EEND,8.99	ent,8.99	DUCK,24.76	dVk,24.76
EENHEID,24.72	enhKt,24.72	UNIT,36.18	junIt,36.18
EERLIJK,172.38	erl@k,172.38	HONEST,72.33	QnIst,72.33
EEUWIG,43.88	ew@x,43.88	ETERNAL,10.73	It3nP,10.73
EIGEN,355.96	KG@,355.96	OWN,459.2	5n,459.2
EIK,1.53	Kk,1.53	OAK,5.61	5k,5.61
EILAND,51.59	KlAnt,51.59	ISLAND,39.57	2l@nd,39.57
EINDE,112.94	Knd@,112.94	END,265.86	End,265.86
ENGEL,32.63	EN@l,32.63	ANGEL,78.27	1n_@l,78.27
ERG,749.2	Erx,749.2	TERRIBLE,94.02	tEr@bP,94.02
ERTS,0.71	Erts,0.71	ORE,1.65	$R,1.65
EZEL,11.41	ez@l,11.41	DONKEY,5.35	dQNkI,5.35
FABEL,0.53	fab@l,0.53	FABLE,0.39	f1bP,0.39
FASE,15.55	faz@,15.55	PHASE,12.33	f1z,12.33

FEE,5.67	fe,5.67	FAIRY,16.69	f8rI,16.69
FEL,6.27	fEl,6.27	FIERCE,4.78	f7s,4.78
FIGUUR,12.19	fiGyr,12.19	FIGURE,129.37	fIg@R,129.37
FILM,174.37	fIlm,174.37	MOVIE,122.96	muvI,122.96
FIRMA,8.78	fIrma,8.78	FIRM,35.27	f3m,35.27
FLAUW,21.5	flMw,21.5	FAINT,9.27	f1nt,9.27
FLITS,3.32	flIts,3.32	FLASH,15.35	fl{S,15.35
FLUIT,7.39	flLt,7.39	WHISTLE,15.45	wIsP,15.45
FLUWEEL,1.12	flywel,1.12	VELVET,6.27	vElvIt,6.27
FOLIE,1.03	foli,1.03	FOIL,1.22	f4l,1.22
FONTEIN,3.96	fOntKn,3.96	FOUNTAIN,6.9	f6ntIn,6.9
FOREL,1.9	forEl,1.9	TROUT,4.02	tr6t,4.02
FORTUIN,14.27	fOrtLn,14.27	FORTUNE,29.76	f$Jun,29.76
FRASE,0.09	fraz@,0.09	PHRASE,9.1	fr1z,9.1
FUNCTIE,10.15	f}Nksi,10.15	FUNCTION,11.08	fVNkSH,11.08
GASTHEER,7.55	xAsther,7.55	HOST,15.02	h5st,15.02
GAT,49.97	xAt,49.97	HOLE,58.22	h5l,58.22
GAZON,3.09	xazOn,3.09	LAWN,12.35	l$n,12.35
GEBED,10.59	x@bEt,10.59	PRAYER,15.78	pr1@R,15.78
GEBOUW,68.26	x@bMw,68.26	BUILDING,99.57	bIldIN,99.57
GEBREK,14.02	x@brEk,14.02	LACK,17.75	l{k,17.75
GEDURFD,1.39	x@d}rft,1.39	DARING,4.2	d8rIN,4.2
GEEL,9.56	xel,9.56	YELLOW,33.8	jEl5,33.8
GEEST,94.97	xest,94.97	SPIRIT,49.35	spIrIt,49.35
GEEUW,0.25	xew,0.25	YAWN,1	j$n,1
GEHEIM,101.76	x@hKm,101.76	SECRET,109.51	sikrIt,109.51
GEHOOR,5.95	x@hor,5.95	HEARING,44.73	h7rIN,44.73
GEIT,8.07	xKt,8.07	GOAT,10.53	g5t,10.53
GELD,793.61	xElt,793.61	MONEY,640.76	mVnI,640.76
GELDEND,3	xEld@nt,3	GOING,404	g5IN,404
GELEI,0.71	Z@lK,0.71	JELLY,7.12	_ElI,7.12
GELOOF,395.2	x@lof,395.2	FAITH,46.33	f1T,46.33
GEMENGD,3.11	Z@ni,3.11	MIXED,21.37	mIkst,21.37
GENIE,25.22	Z@ni,25.22	GENIUS,34.76	_inj@s,34.76
GERING,0.96	x@rIN,0.96	MINOR,12.82	m2n@R,12.82
GERST,0.5	xErst,0.5	BARLEY,0.9	b#lI,0.9
GESCHENK,18.75	x@sxENk,18.75	GIFT,64.51	gIft,64.51
GESLACHT,8.1	x@slAxt,8.1	GENDER,2.8	_End@R,2.8
GESP,0.96	xEsp,0.96	BUCKLE,5.04	bVkP,5.04

GETIJ,1.35	x@tK,1.35	TIDE,7.35	t2d,7.35
GEUL,1.21	x|l,1.21	TRENCH,3.08	trEnJ,3.08
GEVAAR,91.54	x@var,91.54	DANGER,43.67	d1n_@R,43.67
GEVAL,137.21	x@vAl,137.21	CASE,282.41	k1s,282.41
GEVANG,1.14	x@vAN,1.14	PRISON,66.04	prIzH,66.04
GEVECHT,48.98	x@vExt,48.98	FIGHT,201.08	f2t,201.08
GEVOEL,128.27	x@vul,128.27	FEELING,51.79	filIN,51.79
GEVOLG,7.18	x@vOlx,7.18	RESULT,19.76	rIzVlt,19.76
GEWEER,55.39	x@wer,55.39	RIFLE,14.57	r2fP,14.57
GEWELF,0.27	x@wElf,0.27	VAULT,12.04	v$lt,12.04
GEWICHT,17.29	x@wIxt,17.29	WEIGHT,36.27	w1t,36.27
GEWIN,0.87	x@wIn,0.87	GAIN,13.73	g1n,13.73
GEZICHT,183.63	x@zIxt,183.63	FACE,289.16	f1s,289.16
GEZOEM,0.5	x@zum,0.5	BUZZ,15.96	bVz,15.96
GEZOND,31.83	x@zOnt,31.83	HEALTHY,24.75	hElTI,24.75
GIDS,11.39	xIts,11.39	GUIDE,17.84	g2d,17.84
GIF,13.56	xIf,13.56	POISON,24.55	p4zH,24.55
GIPS,2.36	xIps,2.36	PLASTER,2.63	pl#st@R,2.63
GLAD,8	xlAt,8	SMOOTH,18.27	smuD,18.27
GLAS,57.15	xlAs,57.15	GLASS,60.71	gl#s,60.71
GLAZUUR,1.42	xlazyr,1.42	ICING,1.39	2sIN,1.39
GLETSJER,1.44	glEtS@r,1.44	GLACIER,0.75	gl{sj@R,0.75
GLIMLACH,14.43	xlImlAx,14.43	SMILE,58	sm2l,58
GLOED,2.08	xlut,2.08	GLOW,5.75	gl5,5.75
GOLF,17.54	xOlf,17.54	WAVE,21.25	w1v,21.25
GOUD,61.99	xMt,61.99	GOLD,78.94	g5ld,78.94
GRAAD,4.02	xrat,4.02	DEGREE,14.88	dIgri,14.88
GRAF,34.53	xrAf,34.53	GRAVE,26.27	gr1v,26.27
GRAFIEK,0.98	xrafik,0.98	GRAPH,0.75	gr#f,0.75
GRAP,51.02	xrAp,51.02	JOKE,73.02	_5k,73.02
GRAPPIG,128.45	xrAp@x,128.45	FUNNY,218.18	fVnI,218.18
GRAS,18.75	xrAs,18.75	GRASS,16.78	gr#s,16.78
GRATIS,45.37	xrat@s,45.37	FREE,177.53	fri,177.53
GRENS,37.96	xrEns,37.96	FRONTIER,3.27	frVnt7R,3.27
GRETIG,1.99	xret@x,1.99	EAGER,6.86	ig@R,6.86
GROEN,28.06	xrun,28.06	GREEN,72.47	grin,72.47
GROND,110.25	xrOnt,110.25	GROUND,72.47	gr6nd,72.47
GROOT,237.51	xrot,237.51	LARGE,41.45	l#_,41.45
GROOTTE,7.71	xrot@,7.71	SIZE,46.14	s2z,46.14

GULDEN,1.92	x}ld@,1.92	GOLDEN,23.27	g5ld@n,23.27
HAAI,9.44	haj,9.44	SHARK,14.98	S#k,14.98
HAAN,4.28	han,4.28	ROOSTER,3.86	rust@R,3.86
HAAR,43.3	har,43.3	HAIR,43.3	h8R,43.3
HAAS,2.17	has,2.17	HARE,3.82	h8R,3.82
HAGEL,1.35	haG@l,1.35	HAIL,12.02	h1l,12.02
HALS,7.5	hAls,7.5	NECK,59.51	nEk,59.51
HAMER,8.55	ham@r,8.55	HAMMER,12.47	h{m@R,12.47
HANDEL,26.05	hAnd@l,26.05	TRADE,35.2	tr1d,35.2
HANDSCHOEN,6.68	hAntsxun,6.68	GLOVE,10.1	glVv,10.1
HANDTAS,4.32	hAntAs,4.32	HANDBAG,2.63	h{ndb{g,2.63
HANDVAT,2.01	hAntfAt,2.01	HANDLE,108.41	h{ndP,108.41
HAUTAIN,0.18	hot),0.18	HAUGHTY,0.41	h$tI,0.41
HAVER,0.89	hav@r,0.89	OATS,2.06	5ts,2.06
HAVIK,2.36	havIk,2.36	HAWK,12.75	h$k,12.75
HEELAL,11.05	helAl,11.05	UNIVERSE,25.27	junIv3s,25.27
HEET,338.51	het,338.51	HOT,189.84	hQt,189.84
HEILIG,12.21	hKl@x,12.21	HOLY,68.14	h5lI,68.14
HEK,22.78	hEk,22.78	FENCE,16.06	fEns,16.06
HELD,59.64	hElt,59.64	HERO,49.84	h7r5,49.84
HEMEL,116.86	hem@l,116.86	HEAVEN,56.61	hEvH,56.61
HENDEL,3.68	hEnd@l,3.68	LEVER,3.2	liv@R,3.2
HENGST,4.78	hENst,4.78	STALLION,3.2	st{lj@n,3.2
HERFST,9.08	hErfst,9.08	AUTUMN,3.78	$t@m,3.78
HEVIG,4.02	hev@x,4.02	VIOLENT,16.67	v2@l@nt,16.67
HIEL,0.85	hil,0.85	HEEL,7.39	hil,7.39
HOED,35.95	hut,35.95	HAT,64.18	h{t,64.18
HOEK,49.92	huk,49.92	CORNER,52.53	k$n@R,52.53
HOEPEL,7.36	hup@l,7.36	HOOP,2.69	hup,2.69
HOEST,2.88	hust,2.88	COUGH,8.78	kQf,8.78
HOEVE,0.73	huv@,0.73	FARM,30.04	f#m,30.04
HOF,23.19	hOf,23.19	YARD,25.06	j#d,25.06
HOMP,0.27	hOmp,0.27	CHUNK,4.14	JVNk,4.14
HOND,168.65	hOnt,168.65	DOG,192.84	dQg,192.84
HONING,7.16	honIN,7.16	HONEY,300.49	hVnI,300.49
HOOFD,274.05	hoft,274.05	HEAD,371.51	hEd,371.51
HOOG,63.23	hox,63.23	TALL,32.33	t$l,32.33
HOOIBERG,1.76	hojbErx,1.76	HAYSTACK,1.37	h1st{k,1.37
HOORN,8	horn,8	HORN,21.08	h$n,21.08

HOUDING,20.08	hMdIN,20.08	ATTITUDE,26.08	{tItjud,26.08
HUID,39.56	hLt,39.56	SKIN,44.04	skIn,44.04
HUIS,818.9	hLs,818.9	HOUSE,514	h6s,514
HUISDIER,5.83	hLzdir,5.83	PET,20.18	pEt,20.18
HYMNE,0.39	hImn@,0.39	HYMN,1.63	hIm,1.63
IDOOL,1.97	idol,1.97	IDOL,2.76	2dP,2.76
IEP,0.27	ip,0.27	ELM,1.43	Elm,1.43
IJZER,7.11	Kz@r,7.11	IRON,17.94	2@n,17.94
IMPULS,2.4	Imp}ls,2.4	IMPULSE,5.27	ImpVls,5.27
INGANG,16.35	INGAN,16.35	ENTRY,12.16	EntrI,12.16
INHOUD,7.71	InhMt,7.71	CONTENT,7.63	kQntEnt,7.63
INSEKT,0.14	InsEkt,0.14	INSECT,3.16	InsEkt,3.16
IVOOR,1.33	ivor,1.33	IVORY,1.39	2v@rI,1.39
JAAR,762.67	jar,762.67	YEAR,277.92	j7R,277.92
JACHT,23.99	jAxt,23.99	YACHT,8.22	jQt,8.22
JAGER,11.18	jaG@r,11.18	HUNTER,18.35	hVnt@R,18.35
JONG,91.61	jON,91.61	YOUNG,243.18	jVN,243.18
JONGEN,435.7	jON@,435.7	BOY,529.82	b4,529.82
JURK,55.75	j}rk,55.75	DRESS,87.2	drEs,87.2
KAARS,5.88	kars,5.88	CANDLE,8.02	k{ndP,8.02
KAART,79.67	kart,79.67	MAP,31.82	m{p,31.82
KAAS,22.85	kas,22.85	CHEESE,39.04	Jiz,39.04
KABEL,9.76	kab@l,9.76	CABLE,21.73	k1bP,21.73
KACHEL,2.86	kAx@l,2.86	STOVE,7.59	st5v,7.59
KALK,1.01	kAlk,1.01	LIME,3.29	l2m,3.29
KALKOEN,10.4	kAlkun,10.4	TURKEY,22.61	t3kI,22.61
KAM,5.28	kAm,5.28	COMB,6.06	k5m,6.06
KANAAL,13.95	kanal,13.95	CHANNEL,24.41	J{nP,24.41
KANON,6.17	kanOn,6.17	CANNON,8.71	k{n@n,8.71
KANS,264.38	kans,264.38	CHANCE,241.24	J#ns,241.24
KANTOOR,124.42	kAntor,124.42	OFFICE,203.9	QfIs,203.9
KAP,12.28	kAp,12.28	HOOD,15.39	hUd,15.39
KAR,6.75	kAr,6.75	CART,9.04	k#t,9.04
KAST,30.05	kAst,30.05	HIVE,0.98	h2v,0.98
KAT,52.85	kAt,52.85	CAT,66.33	k{t,66.33
KATOEN,5.05	katun,5.05	COTTON,14.18	kQtH,14.18
KAUWGOM,5.85	kMwGOm,5.85	GUM,13.39	gVm,13.39
KEEL,27.28	kel,27.28	THROAT,36.02	Tr5t,36.02
KEIZER,27.33	kKz@r,27.33	EMPEROR,13.53	Emp@r@R,13.53

KERN,9.26	kErn,9.26	NUCLEUS,0.88	njukl7s,0.88
KETTING,19.14	kEtIN,19.14	NECKLACE,9.75	nEklIs,9.75
KEU,1.65	k|,1.65	CUE,7.78	kju,7.78
KIND,333.3	kInt,333.3	CHILD,157.65	J2ld,157.65
KIP,37.89	kIp,37.89	CHICKEN,61.73	JIkIn,61.73
KLAAR,556.28	klar,556.28	READY,387.8	rEdI,387.8
KLANT,35.33	klAnt,35.33	CUSTOMER,15.2	kVst@m@R,15.2
KLAP,29.27	klAp,29.27	SLAP,12.47	sl{p,12.47
KLAUW,3.27	klMw,3.27	CLAW,4.37	kl$,4.37
KLERK,1.62	klErk,1.62	CLERK,12.9	kl#k,12.9
KLIK,3.04	klIk,3.04	CLICK,6.41	klIk,6.41
KLIMAAT,4.12	klimat,4.12	CLIMATE,3.53	kl2mIt,3.53
KLINK,6.72	klINk,6.72	LATCH,1.92	l{J,1.92
KLOK,23.9	klOk,23.9	CLOCK,58.63	klQk,58.63
KNAL,14.82	knAl,14.82	BANG,19.98	b{N,19.98
KNECHT,2.79	knExt,2.79	SERVANT,12.14	s3v@nt,12.14
KNIE,10.24	kni,10.24	KNEE,14.69	ni,14.69
KNOFLOOK,4.39	knOflok,4.39	GARLIC,6	g#lIk,6
KNOOP,9.51	knop,9.51	KNOT,3.69	nQt,3.69
KNOP,16.46	knOp,16.46	BUTTON,28.25	bVtH,28.25
KNUPPEL,7.32	kn}p@l,7.32	BAT,20.63	b{t,20.63
KOE,18.55	ku,18.55	COW,25.51	k6,25.51
KOFFER,33.87	kOf@r,33.87	SUITCASE,13.39	sutk1s,13.39
KOFFIE,133.41	kOfi,133.41	COFFEE,144.53	kQfI,144.53
KOKEN,33.8	kok@,33.8	COOK,45.57	kUk,45.57
KOM,2612.54	kOm,2612.54	SOCKET,1.61	sQkIt,1.61
KOMPAS,4.55	kOmpAs,4.55	COMPASS,4.06	kVmp@s,4.06
KONIJN,18.87	konKn,18.87	RABBIT,20.94	r{bIt,20.94
KONING,138.53	konIN,138.53	KING,129.25	kIN,129.25
KOOPJE,3.38	kopj@,3.38	BARGAIN,12	b#gIn,12
KOOPWAAR,1.42	kopwar,1.42	GOODS,12.31	gUdz,12.31
KOOR,6.88	kor,6.88	CHOIR,5.31	kw2@R,5.31
KOORD,2.26	kort,2.26	CORD,7.02	k$d,7.02
KOORTS,14.77	korts,14.77	FEVER,19.94	fiv@R,19.94
KOPEN,130.64	kop@,130.64	BUY,192.43	b2,192.43
KOPIE,16.85	kopi,16.85	COPY,52.27	kQpI,52.27
KOPPEL,6.95	kOp@l,6.95	COUPLE,223.41	kVpP,223.41
KOPPIG,10.75	kOp@x,10.75	STUBBORN,10.86	stVb@n,10.86
KORAAL,0.82	koral,0.82	CORAL,2.37	kQr@l,2.37

KORREL,0.73	kOr@l,0.73	GRAIN,4.76	gr1n,4.76
KORT,47.52	kOrt,47.52	SHORT,85.63	S$t,85.63
KOST,80.4	kOst,80.4	LIVING,156.53	lIvIN,156.53
KOSTBAAR,5.88	kOzdbar,5.88	VALUABLE,15.59	v{lj9bP',15.59
KOUD,95.95	kMt,95.95	COLD,130.16	k5ld,130.16
KRAAL,1.12	kral,1.12	BEAD,1.12	bid,1.12
KRAAN,6.4	kran,6.4	FAUCET,1.43	f$sIt,1.43
KRACHTIG,10.31	krAxt@x,10.31	POWERFUL,35.12	p6@fUl,35.12
KRAMP,3.27	krAmp,3.27	CRAMP,2.8	kr{mp,2.8
KREEFT,5.53	kreft,5.53	LOBSTER,7.33	lQbst@R,7.33
KREEK,0.96	krek,0.96	CREEK,8.9	krik,8.9
KRIJGER,8.1	krKG@r,8.1	WARRIOR,10.12	wQr7R,10.12
KRIJT,11.82	krKt,11.82	CHALK,3.59	J$k,3.59
KRIK,1.1	krIk,1.1	JACK,251.59	_{k,251.59
KRISTAL,4.87	krIstAl,4.87	CRYSTAL,16.14	krIstP,16.14
KRUISING,4.41	krLsIN,4.41	CROSSING,8.71	krQsIN,8.71
KRUK,2.56	kr}k,2.56	STOOL,3.51	stul,3.51
KRUL,0.3	kr}l,0.3	CURL,2.35	k3l,2.35
KUDDE,12.42	k}d@,12.42	HERD,7.06	h3d,7.06
KUIL,3	kLl,3	PIT,13.22	pIt,13.22
KUIP,0.85	kLp,0.85	TUB,12.69	tVb,12.69
KUNST,37.09	k}nst,37.09	ART,70.8	#t,70.8
KURK,1.62	k}rk,1.62	CORK,2.86	k$k,2.86
KUST,31.88	k}st,31.88	COAST,26.69	k5st,26.69
KWARTAAL,1.51	kwArtal,1.51	QUARTER,26.02	kw$t@R,26.02
LAAG,31.1	lax,31.1	LOW,59.14	l5,59.14
LAARS,4.18	lars,4.18	BOOT,11.14	but,11.14
LAATST,60.49	latst,60.49	LAST,723.1	l#st,723.1
LAFAARD,20.56	lAfart,20.56	COWARD,14.39	k6@d,14.39
LANG,596.85	lAN,596.85	LONG,675.16	lQN,675.16
LAST,53.49	lAst,53.49	BURDEN,9.82	b3dH,9.82
LASTIG,67.69	lAst@x,67.69	ANNOYING,11.12	@n4IN,11.12
LAWAAI,17.08	lawaj,17.08	NOISE,34.88	n4z,34.88
LEEFTIJD,44.98	leftKt,44.98	AGE,79.2	1_,79.2
LEEG,61.74	lex,61.74	EMPTY,47.24	EmptI,47.24
LEERLING,12.99	lerlIN,12.99	PUPIL,3.14	pjupP,3.14
LEGAAL,11.75	leGal,11.75	LEGAL,35.71	ligP,35.71
LEGER,107.98	leG@r,107.98	ARMY,85.69	#mI,85.69
LEK,10.45	lEk,10.45	LEAK,10.14	lik,10.14

LELIE,0.85	leli,0.85	LILY,26.86	lIlI,26.86
LEPEL,5.01	lep@l,5.01	SPOON,7.61	spun,7.61
LERAAR,29.66	lerar,29.66	TEACHER,55.73	tiJ@R,55.73
LETSEL,3.11	lEts@l,3.11	INJURY,10.2	In_@rI,10.2
LEVEND,55.61	lev@nt,55.61	ALIVE,154.47	@l2v,154.47
LEZER,1.53	lez@r,1.53	READER,5.45	rid@R,5.45
LICHAAM,147.54	lIxam,147.54	BODY,195.53	bQdI,195.53
LICHTER,4.41	lIxt@r,4.41	LIGHTER,8.96	l2t@R,8.96
LID,33.18	lIt,33.18	MEMBER,28.78	mEmb@R,28.78
LIED,21.27	lit,21.27	SONG,93.69	sQN,93.69
LINKS,68.56	lINks,68.56	LEFT,484.45	lEft,484.45
LINT,7.98	lInt,7.98	RIBBON,5.06	rIb@n,5.06
LITTEKEN,9.81	lItek@,9.81	SCAR,8.47	sk#R,8.47
LOGE,2.04	l<z@,2.04	LODGE,6.69	lQ_,6.69
LOOPBAAN,2.74	loban,2.74	CAREER,45.2	k@r7R,45.2
LOPER,4.34	lop@r,4.34	RUNNER,4.96	rVn@R,4.96
LUCHT,89	l}xt,89	SKY,44.8	sk2,44.8
LUS,1.37	l}s,1.37	TAB,5.76	t{b,5.76
MAAG,23.55	max,23.55	STOMACH,33.82	stVm@k,33.82
MAALTIJD,15.37	maltKt,15.37	MEAL,28.86	mil,28.86
MAAN,42.1	man,42.1	MOON,49.96	mun,49.96
MAAT,69.18	mat,69.18	PAL,57.59	p{l,57.59
MAJOOR,52.8	major,52.8	MAJOR,104.76	m1_@R,104.76
MAMA,206.45	mAma,206.45	MOM,430.39	mQm,430.39
MAND,4.3	mAnt,4.3	BASKET,13.18	b#skIt,13.18
MANIER,257.4	manir,257.4	WAY,1424.73	w1,1424.73
MAP,4.53	mAp,4.53	FOLDER,1.63	f5ld@R,1.63
MARKER,0.23	mArk@r,0.23	MARKER,5.24	m#k@R,5.24
MARMER,1.97	mArm@r,1.97	MARBLE,5.22	m#bP,5.22
MATROOS,3.98	mAtros,3.98	SAILOR,12.39	s1l@R,12.39
MEDISCH,12.23	medis,12.23	MEDICAL,54.39	mEdIkP,54.39
MEEUW,0.64	mew,0.64	GULL,1.1	gVl,1.1
MEID,114.52	mKt,114.52	MAID,22.82	m1d,22.82
MEISJE,382.74	mKsj@,382.74	GIRL,557.12	g3l,557.12
MENEER,518.44	m@ner,518.44	MISTER,45.61	mIst@R,45.61
MENING,34.48	menIN,34.48	OPINION,42	@pInj@n,42
MENS,144.66	mEns,144.66	HUMAN,124.76	hjum@n,124.76
MENTAAL,3.84	mEntal,3.84	MENTAL,19.65	mEntP,19.65
MERG,0.69	mErx,0.69	MARROW,2.96	m{r5,2.96

MERK,19.12	mErk,19.12	MARK,82.02	m#k,82.02
MERRIE,2.72	mEri,2.72	MARE,2.9	m8R,2.9
MES,46.24	mEs,46.24	KNIFE,46.8	n2f,46.8
METAAL,9.54	metal,9.54	METAL,19.45	mEtP,19.45
METHODE,8.48	metod@,8.48	METHOD,7.88	mit@R,7.88
METRO,13.1	metro,13.1	SUBWAY,10.71	sVbw1,10.71
MIDDEL,14.5	mId@l,14.5	MEANS,218.35	minz,218.35
MIDDEN,47.29	mId@,47.29	MIDDLE,89.2	mIdP,89.2
MIER,2.54	mir,2.54	ANT,5.35	{nt,5.35
MIJT,0.07	mKt,0.07	STACK,6.1	st{k,6.1
MINNAAR,12.87	mInar,12.87	LOVER,26.63	lVv@R,26.63
MINUUT,42.69	minyt,42.69	MINUTE,377.49	mInIt,377.49
MIS,219.69	mIs,219.69	MASS,17.25	m{s,17.25
MISDAAD,43.22	mIzdat,43.22	CRIME,71.24	kr2m,71.24
MISSIE,53.4	mIsi,53.4	MISSION,47.06	mISH,47.06
MIST,45.6	mIst,45.6	FOG,9.45	fQg,9.45
MODDER,12.81	mOd@r,12.81	MUD,14.82	mVd,14.82
MOE,89.94	mu,89.94	TIRED,112.65	t2@d,112.65
MOEDER,652.49	mud@r,652.49	MOTHER,479.92	mVD@R,479.92
MOEDIG,18.84	mud@x,18.84	BRAVE,31.71	br1v,31.71
MOEITE,68.6	mujt@,68.6	TROUBLE,223.55	trVbP,223.55
MOER,9.22	mur,9.22	NUT,15.63	nVt,15.63
MOL,7.55	mOl,7.55	MOLE,8.06	m5l,8.06
MOLEN,4.53	mol@,4.53	MILL,9.53	mIl,9.53
MOND,165.93	mOnt,165.93	MOUTH,104.41	m6T,104.41
MONNIK,5.95	mOn@k,5.95	MONK,7.37	mVNk,7.37
MONSTER,49.97	mOnst@r,49.97	SAMPLE,14.59	s#mpP,14.59
MORAAL,6.88	moral,6.88	MORAL,13.51	mQr@l,13.51
MORGEN,423.61	mOrG@,423.61	TOMORROW,335.96	t@mQr5,335.96
MOS,1.14	mOs,1.14	MOSS,2.84	mQs,2.84
MOSTERD,4.99	mOst@rt,4.99	MUSTARD,6.45	mVst@d,6.45
MOTIEF,15.5	motif,15.5	MOTIVE,13.24	m5tIv,13.24
MOUT,0.11	mMt,0.11	MALT,1.65	m$lt,1.65
MOUW,4.18	mMw,4.18	SLEEVE,5.61	sliv,5.61
MUF,0.57	m}f,0.57	STALE,2.92	st1l,2.92
MUIS,11.14	mLs,11.14	MOUSE,19.12	m6s,19.12
MUNT,10.89	m}nt,10.89	MINT,5.43	mInt,5.43
MUUR,66.89	myr,66.89	WALL,70.69	w$l,70.69
MUZIEK,107.46	myzik,107.46	MUSIC,151.65	mjuzIk,151.65

NAAD,1.78	nat,1.78	SEAM,0.65	sim,0.65
NAAM,470.6	nam,470.6	NAME,641.86	n1m,641.86
NACHT,204.44	nAxt,204.44	NIGHT,866.04	n2t,866.04
NAGEL,4.05	naG@l,4.05	NAIL,18.65	n1l,18.65
NATUUR,26.96	natyr,26.96	NATURE,45.16	n1J@R,45.16
NEEF,48.48	nef,48.48	NEPHEW,16.59	nEvju,16.59
NEGEN,64.28	neG@,64.28	NINE,67.47	n2n,67.47
NERVEUS,42.9	nErv|s,42.9	NERVOUS,67.16	n3v@s,67.16
NETTO,0.64	nEto,0.64	NET,15.55	nEt,15.55
NEUTRAAL,4.18	n|tral,4.18	NEUTRAL,4.22	njutr@l,4.22
NICHT,22.96	nIxt,22.96	NIECE,9.53	nis,9.53
NIER,3.91	nir,3.91	KIDNEY,9.69	kIdnI,9.69
NIETIG,2.77	nit@x,2.77	VOID,4.1	v4d,4.1
NIEUWS,178.42	niws,178.42	NEWS,164.69	njuz,164.69
NIKKEL,0.39	nIk@l,0.39	NICKEL,8.45	nIkP,8.45
NOORD,5.76	nort,5.76	NORTH,63.88	n$T,63.88
NUL,17.61	n}l,17.61	ZERO,21.45	z7r5,21.45
OCHTEND,39.54	Oxt@nt,39.54	MORNING,439	m$nIN,439
OMA,72.86	oma,72.86	GRANDMA,45.59	gr{nm#,45.59
ONDEUGD,1.12	Ond|xt,1.12	VICE,18.63	v2s,18.63
ONDIEP,1.23	Ondip,1.23	SHALLOW,5.9	S{l5,5.9
ONKRUID,3.18	ONkrLt,3.18	WEED,11.76	wid,11.76
ONTZET,1.33	OntsEt,1.33	RELIEF,14.59	rIlif,14.59
ONZIN,111.23	OnzIn,111.23	NONSENSE,28.47	nQns@ns,28.47
OOG,68.4	ox,68.4	EYE,111.78	2,111.78
OOM,108.78	om,108.78	UNCLE,124.06	VNkP,124.06
OOR,25.02	or,25.02	EAR,32	7R,32
OORLOG,178.96	orlOx,178.96	WAR,174.75	w$R,174.75
OPRIT,3.93	OprIt,3.93	DRIVEWAY,6.73	dr2vw1,6.73
OPROEP,11.07	Oprup,11.07	CALL,861.39	k$l,861.39
OPSLAG,8.44	OpslAx,8.44	STORAGE,8.86	st$rI_,8.86
OPTIE,11.21	Opsi,11.21	OPTION,14.43	QpSH,14.43
OUD,183.26	Mt,183.26	OLD,608.94	5ld,608.94
PAARD,83.63	part,83.63	HORSE,92.88	h$s,92.88
PAD,41.99	pAt,41.99	TOAD,5.69	t5d,5.69
PAK,313.2	pAk,313.2	SUIT,68.61	sut,68.61
PANEEL,2.01	panel,2.01	PANEL,7.29	p{nP,7.29
PANIEK,39.86	panik,39.86	PANIC,21.84	p{nIk,21.84
PANISCH,0.3	panis,0.3	FRANTIC,2.27	fr{ntIk,2.27

PANTER,1.53	pAnt@r,1.53	PANTHER,2.57	p{nT@R,2.57
PAPIER,31.05	papir,31.05	PAPER,103.35	p1p@R,103.35
PAREL,3.02	par@l,3.02	PEARL,15.67	p3l,15.67
PARFUM,10.89	pArf}m,10.89	PERFUME,11.43	p3fjum,11.43
PARTIJ,22.62	pArtK,22.62	PARTY,233.14	p#tI,233.14
PASSIEF,1.14	pAsif,1.14	PASSIVE,2.18	p{sIv,2.18
PASTA,5.92	pAsta,5.92	PASTE,1.71	p1st,1.71
PAUS,12.62	pMs,12.62	POPE,10.71	p5p,10.71
PAUZE,27.19	pMz@,27.19	PAUSE,5.39	p$z,5.39
PEPER,3.8	pep@r,3.8	PEPPER,8.8	pEp@R,8.8
PERZIK,2.29	pErzIk,2.29	PEACH,6.35	piJ,6.35
PIJN,266.16	pKn,266.16	PAIN,97.94	p1n,97.94
PIJP,13.81	pKp,13.81	PIPE,19.39	p2p,19.39
PINDA,2.13	pInda,2.13	PEANUT,12.35	pinVt,12.35
PIRAAT,6.4	pirat,6.4	PIRATE,7.35	p2@r@t,7.35
PISTOOL,102.63	pistol,102.63	PISTOL,10.06	pIstP,10.06
PIT,5.53	pIt,5.53	WICK,2.51	wIk,2.51
PITTIG,3.75	pIt@x,3.75	SPICY,3.31	sp2sI,3.31
PLAFOND,8.58	plafOnt,8.58	CEILING,8.35	silIN,8.35
PLAK,3.13	plAk,3.13	SLICE,8.53	sl2s,8.53
PLANEET,46.33	planet,46.33	PLANET,38.73	pl{nIt,38.73
PLANK,11.21	plANk,11.21	SHELF,6.96	SElf,6.96
PLEK,177.41	plEk,177.41	SPOT,61.57	spQt,61.57
PLICHT,28.1	plIxt,28.1	DUTY,50.96	djutI,50.96
PLOEG,9.81	plux,9.81	SHIFT,22.82	SIft,22.82
PLONS,0.48	plOns,0.48	SPLASH,4.22	spl{S,4.22
PLUCHE,0.32	plyS,0.32	PLUSH,0.55	plVS,0.55
POEDER,6.15	pud@r,6.15	POWDER,16.04	p6d@R,16.04
POEF,1.01	puf,1.01	CUSHION,2.16	kUSH,2.16
POGING,22.57	poGIN,22.57	ATTEMPT,19.12	@tEmpt,19.12
POLITIE,346.17	politsi,346.17	POLICE,236.16	p@lis,236.16
POMPOEN,2.49	pOmpun,2.49	PUMPKIN,10.84	pVmpkIn,10.84
POORT,25.06	port,25.06	GATE,32.04	g1t,32.04
POOT,6.95	pot,6.95	PAW,3.12	p$,3.12
PORTIE,4.62	pOrsi,4.62	PORTION,4.33	p$SH,4.33
POT,30.62	pOt,30.62	JAR,8.31	_#R,8.31
POTLOOD,5.44	pOtlot,5.44	PENCIL,9.86	pEnsP,9.86
PRIJS,86.6	prKs,86.6	PRICE,53.37	pr2s,53.37
PRIK,4.6	prIk,4.6	PRICK,14.12	prIk,14.12

PRIMAAT,0.3	primat,0.3	PRIMATE,0.69	pr2m1t,0.69
PRINS,45.26	prIns,45.26	PRINCE,45.08	prIns,45.08
PROBLEEM,341.78	problem,341.78	PROBLEM,330.06	prQbl@m,330.06
PROCENT,44.52	prosEnt,44.52	PERCENT,25.75	p@sEnt,25.75
PRODUKT,0.55	prod}kt,0.55	PRODUCT,14.75	prQdVkt,14.75
PROOI,7.82	proj,7.82	PREY,5.51	pr1,5.51
PROZA,0.39	proza,0.39	PROSE,0.94	pr5z,0.94
PRUIM,1.12	prLm,1.12	PLUM,3.41	plVm,3.41
PUBLIEK,41.87	pyblik,41.87	PUBLIC,71.08	pVblIk,71.08
PUZZEL,5.26	p}z@l,5.26	PUZZLE,7.33	pVzP,7.33
RAAM,70.84	ram,70.84	WINDOW,86	wInd5,86
RADIJS,0.41	radKs,0.41	RADISH,0.61	r{dIS,0.61
RAIL,0.46	rel,0.46	RAIL,4.57	r1l,4.57
RAKET,14.82	rakEt,14.82	ROCKET,11.84	rQkIt,11.84
RAMP,25.89	rAmp,25.89	DISASTER,17.27	dIz#st@R,17.27
RAND,19.12	rAnt,19.12	EDGE,23.51	E_,23.51
RAUW,4.67	rMw,4.67	RAW,10.18	r$,10.18
RECHTER,63.28	rExt@r,63.28	JUDGE,79.67	_V_,79.67
REDEN,163.67	red@,163.67	REASON,193.29	rizH,193.29
REFREIN,1.14	r@frKn,1.14	CHORUS,6.08	k$r@s,6.08
REGEN,26.48	reG@,26.48	RAIN,48.9	r1n,48.9
REIKEN,2.74	rKk@,2.74	RANGE,22.76	r1n_,22.76
REIS,90.37	rKs,90.37	TRIP,82.39	trIp,82.39
REK,2.38	rEk,2.38	RACK,7.78	r{k,7.78
REL,2.7	rEl,2.7	RIOT,6.49	r2@t,6.49
RESPONS,0.94	rEspOns,0.94	RESPONSE,16.1	rIspQns,16.1
RIEM,14.16	rim,14.16	BELT,24.35	bElt,24.35
RIJ,84.84	rK,84.84	ROW,26.33	r5,26.33
RIJK,79.15	rKk,79.15	RICH,80.39	rIJ,80.39
RIJP,8.1	rKp,8.1	MATURE,8.22	m@tj9R,8.22
RIJST,9.17	rKst,9.17	RICE,15.08	r2s,15.08
RIJWIEL,1	rKwil,1	BICYCLE,6	b2sIkP,6
RIMBOE,1.33	rImbu,1.33	BUSH,14.12	bUS,14.12
RIMPEL,0.8	rImp@l,0.8	WRINKLE,1.88	rINkP,1.88
RIT,14.77	rIt,14.77	RIDE,135.37	r2d,135.37
RITME,8.85	rItm@,8.85	RHYTHM,10.9	rID@m,10.9
RIVIER,52.87	rivir,52.87	RIVER,55.47	rIv@R,55.47
ROEST,2.31	rust,2.31	RUST,2.49	rVst,2.49
ROET,1.46	rut,1.46	SOOT,1.08	sUt,1.08

ROG,1.39	rOx,1.39	RAY,97.37	r1,97.37
ROK,7.23	rOk,7.23	SKIRT,9.96	sk3t,9.96
ROMP,6.47	rOmp,6.47	HULL,4.22	hVl,4.22
RONDE,21.82	rOnd@,21.82	ROUND,66.53	r6nd,66.53
ROOK,46.63	rok,46.63	SMOKE,65.43	sm5k,65.43
ROOM,7.59	rom,7.59	CREAM,48.71	krim,48.71
ROT,87.15	rOt,87.15	ROTTEN,17.47	rQtH,17.47
ROZIJN,0.3	rozKn,0.3	RAISIN,1.63	r1zH,1.63
RUIL,15.39	rLl,15.39	EXCHANGE,20.22	IksJ1n_,20.22
RUIMTE,77.25	rLmt@,77.25	SPACE,66.06	sp1s,66.06
SAAI,31.47	saj,31.47	DULL,12.08	dVl,12.08
SAMEN,370.64	sam@,370.64	TOGETHER,383.39	t@gED@R,383.39
SAP,7.09	sAp,7.09	JUICE,9.9	_us,9.9
SATIJN,0.78	satKn,0.78	SATIN,2.61	s{tIn,2.61
SAUS,9.56	sMs,9.56	SAUCE,15.59	s$s,15.59
SCALP,0.94	skAlp,0.94	SCALP,3.69	sk{lp,3.69
SCHAAP,6.54	sxap,6.54	SHEEP,13.43	Sip,13.43
SCHAAR,6.36	sxar,6.36	SCISSORS,6.69	sIz@z,6.69
SCHAARS,2.22	schaars,2.22	SCARCE,1.71	sk8s,1.71
SCHADE,31.83	sxad@,31.83	HARM,31.78	h#m,31.78
SCHADUW,20.92	sxadyw,20.92	SHADOW,21.18	S{d5,21.18
SCHAKEL,14.64	sxak@l,14.64	LINK,11.94	lINk,11.94
SCHAKEN,4	sxak@,4	CHESS,7.45	JEs,7.45
SCHAT,264.03	sxAt,264.03	TREASURE,19.06	trEZ@R,19.06
SCHATTIG,28.68	sxAt@x,28.68	CUTE,87.75	kjut,87.75
SCHEIDEN,25.29	sxKd@,25.29	SEPARATE,21.71	sEp@r1t,21.71
SCHEMER,0.37	sxem@r,0.37	TWILIGHT,3.06	tw2l2t,3.06
SCHEP,4.62	sxEp,4.62	SHOVEL,6.84	SVvP,6.84
SCHEPSEL,4.41	sxEps@l,4.41	CREATURE,21.41	kriJ@R,21.41
SCHERP,20.83	sxErp,20.83	SHARP,23.78	S#p,23.78
SCHIL,1.12	sxIl,1.12	PEEL,5.35	pil,5.35
SCHILDER,10.5	sxIld@r,10.5	PAINTER,6.75	p1nt@R,6.75
SCHOK,11.66	sxOk,11.66	SHOCK,28.78	SQk,28.78
SCHOON,49.28	sxon,49.28	CLEAN,121.24	klin,121.24
SCHOOT,49.35	sxot,49.35	LAP,13.47	l{p,13.47
SCHOP,16.19	sxOp,16.19	KICK,73.41	kIk,73.41
SCHOUDER,18.57	sxMd@r,18.57	SHOULDER,26.2	S5ld@R,26.2
SCHRIJN,0.16	sxrKn,0.16	SHRINE,2.96	Sr2n,2.96
SCHROEF,1.97	sxruf,1.97	SCREW,37.49	skru,37.49

SCHUIM,3.54	sxLm,3.54	FOAM,3.51	f5m,3.51
SCHULD,178.1	sx}lt,178.1	GUILT,14.9	gIlt,14.9
SCHULDIG,85.91	sx}ld@x,85.91	GUILTY,62.29	gIltI,62.29
SEKS,100.14	sEks,100.14	SEX,152.18	sEks,152.18
SEREEN,0.69	seren,0.69	SERENE,1.24	sIrin,1.24
SERVET,1.78	sErvEt,1.78	NAPKIN,3.61	n{pkIn,3.61
SESSIE,6.15	sEsi,6.15	SESSION,13.29	sESH,13.29
SFINX,0.91	sfINks,0.91	SPHINX,1.02	sfINks,1.02
SHORTS,0.8	SOrts,0.8	SHORTS,9.41	S$ts,9.41
SIGAAR,9.83	siGar,9.83	CIGAR,12.94	sIg#R,12.94
SIMPEL,52.92	sImp@l,52.92	SIMPLE,89.31	sImpP,89.31
SJAAL,5.31	Sal,5.31	SCARF,4.69	sk#f,4.69
SKELET,2.84	sk@lEt,2.84	SKELETON,5.12	skElItH,5.12
SLAAF,17.75	slaf,17.75	SLAVE,18.43	sl1v,18.43
SLAGER,6.63	slaG@r,6.63	BUTCHER,8.51	bUJ@R,8.51
SLAK,2.38	slAk,2.38	SNAIL,1.76	sn1l,1.76
SLANG,21.59	slAN,21.59	SNAKE,22.35	sn1k,22.35
SLANK,3.16	slANk,3.16	SLIM,11.86	slIm,11.86
SLECHT,267.81	slExt,267.81	BAD,545.18	b{d,545.18
SLET,28.58	slEt,28.58	TART,2.39	t#t,2.39
SLEUTEL,80.7	sl|t@l,80.7	KEY,86.86	ki,86.86
SLINGER,1.46	slIN@r,1.46	SLING,2.29	slIN,2.29
SLUW,3.2	slyw,3.2	SLY,2.67	sl2,2.67
SMAAK,29.18	smak,29.18	TASTE,51.31	t1st,51.31
SMARAGD,1.92	smarAxt,1.92	EMERALD,2.57	Em@r@ld,2.57
SMERIG,24.35	smer@x,24.35	FILTHY,16.43	fIlTI,16.43
SMERIS,22.91	smer@s,22.91	COP,86.14	kQp,86.14
SNEL,463.6	snEl,463.6	SWIFT,3.86	swIft,3.86
SNELHEID,24.22	snElhKt,24.22	SPEED,41.25	spid,41.25
SOCIAAL,7.94	soSal,7.94	SOCIAL,33.39	s5SP,33.39
SODA,2.61	soda,2.61	SODA,19.84	s5d@,19.84
SOEP,17.84	sup,17.84	SOUP,25.2	sup,25.2
SOK,3.11	sOk,3.11	SOCK,8.98	sQk,8.98
SOLDAAT,53.03	sOldat,53.03	SOLDIER,38.92	s5l_@R,38.92
SOM,3.89	sOm,3.89	SUM,10.25	sVm,10.25
SOMBER,6.88	sort,6.88	GLOOMY,2.41	glumI,2.41
SOORT,222.02	sort,222.02	KIND,590.69	k2nd,590.69
SOP,0.62	sOp,0.62	SUDS,0.47	sVdz,0.47
SPANNEND,20.76	spAn@nt,20.76	EXCITING,34.82	Iks2tIN,34.82

SPANNING,16.17	spAnIN,16.17	TENSION,8.55	tEnSH,8.55
SPECIAAL,50.63	speSal,50.63	SPECIAL,148.57	spESP,148.57
SPEEKSEL,2.95	speks@l,2.95	SALIVA,2.65	s@l2v@,2.65
SPELEN,247.77	spel@,247.77	PLAY,354.53	pl1,354.53
SPIEGEL,27.44	spiG@l,27.44	MIRROR,24.18	mIr@R,24.18
SPIER,2.7	spir,2.7	MUSCLE,13.61	mVsP,13.61
SPION,18.5	spijOn,18.5	SPY,20.06	sp2,20.06
SPLEET,1.78	splet,1.78	CRACK,32.84	kr{k,32.84
SPOED,2.77	sput,2.77	HURRY,173.65	hVrI,173.65
SPONS,2.49	spOns,2.49	SPONGE,6.71	spVn_,6.71
SPOOK,12.76	spok,12.76	PHANTOM,4.08	f{nt@m,4.08
SPORT,20.12	spOrt,20.12	SPORTS,27.59	sp$ts,27.59
SPUIT,9.79	spLt,9.79	SYRINGE,1.94	sIrIn_,1.94
STAART,17.95	start,17.95	TAIL,23.9	t1l,23.9
STAD,272.61	stAt,272.61	CITY,169.1	sItI,169.1
STAF,11.62	stAf,11.62	STAFF,32	st#f,32
STAKING,3.25	stakIN,3.25	STRIKE,45.57	str2k,45.57
STAM,12.46	stAm,12.46	TRUNK,19.8	trVNk,19.8
STANK,11.14	stANk,11.14	STENCH,2.22	stEnJ,2.22
STAPEL,10.22	stap@l,10.22	PILE,13.18	p2l,13.18
STAR,8.03	stAr,8.03	RIGID,1.86	rI_Id,1.86
STAREN,13.08	star@,13.08	STARE,9.96	st8R,9.96
STEIL,1.76	stKl,1.76	STEEP,2.45	stip,2.45
STEM,86.53	stEm,86.53	VOICE,86.16	v4s,86.16
STER,43.77	stEr,43.77	STAR,81.35	st#R,81.35
STERKTE,8.32	stErkt@,8.32	STRENGTH,36.92	strENT,36.92
STEUN,32.68	st|n,32.68	SUPPORT,0.73	s@p$t,0.73
STIEKEM,11.69	stik@m,11.69	SNEAKY,4.24	snikI,4.24
STIER,9.6	stir,9.6	BULL,27.51	bUl,27.51
STIJF,7.96	stKf,7.96	STIFF,10.25	stIf,10.25
STIJGING,0.82	stKGIN,0.82	RISE,27.43	r2z,27.43
STILTE,37.87	stIlt@,37.87	SILENCE,25.37	s2l@ns,25.37
STOF,29.25	stOf,29.25	MATERIAL,22.14	m@t7r7l,22.14
STOK,14.48	stOk,14.48	STICK,97.12	stIk,97.12
STOM,103.91	stOm,103.91	STUPID,184.41	stjupId,184.41
STOMP,2.33	stOmp,2.33	STUMP,2.45	stVmp,2.45
STOOM,5.88	stom,5.88	STEAM,13.45	stim,13.45
STOPPEL,0.05	stOp@l,0.05	BRISTLE,0.22	brIsP,0.22
STRAAT,100.34	strat,100.34	STREET,148.18	strit,148.18

STRAK,16.76	strAk,16.76	TIGHT,50.92	t2t,50.92
STRENG,13.47	strEN,13.47	STRAND,1.84	str{nd,1.84
STRIJD,44.73	strKt,44.73	STRUGGLE,13.37	strVgP,13.37
STRO,1.78	stro,1.78	STRAW,6.24	str$,6.24
STROOM,29.87	strom,29.87	STREAM,8.04	strim,8.04
STROP,4.37	strOp,4.37	NOOSE,2.18	nus,2.18
STRUIK,2.7	strLk,2.7	SHRUB,0.27	SrVb,0.27
STUK,175.67	st}k,175.67	PIECE,124.49	pis,124.49
SUBTIEL,5.69	s}ptil,5.69	SUBTLE,6.94	sVtP,6.94
SUCCES,103.64	syksEs,103.64	SUCCESS,27.25	s@ksEs,27.25
SUF,4.57	s}f,4.57	DROWSY,0.73	dr6zI,0.73
SYMBOOL,13.84	sImbol,13.84	SYMBOL,8.59	sImbP,8.59
SYSTEEM,45.83	sistem,45.83	SYSTEM,91.51	sIst@m,91.51
TAAI,9.03	taj,9.03	TOUGH,90.51	tVf,90.51
TAAK,42.63	tak,42.63	TASK,12.73	t#sk,12.73
TACTVOL,1.28	tAktfOl,1.28	TACTFUL,0.57	t{ktfUl,0.57
TAFEL,83.4	taf@l,83.4	TABLE,105.63	t1bP,105.63
TALRIJK,0.82	tAlrKk,0.82	NUMEROUS,3.59	njum@r@s,3.59
TAND,8.07	tAnt,8.07	TOOTH,13.57	tuT,13.57
TANG,4.3	tAN,4.3	PLIERS,1.16	pl2@z,1.16
TANTE,62.34	tAnt@,62.34	AUNT,55.2	#nt,55.2
TAPIJT,10.77	tapKt,10.77	CARPET,11.65	k#pIt,11.65
TAS,58.36	tAs,58.36	BAG,94.04	b{g,94.04
TEDER,2.77	ted@r,2.77	TENDER,8.88	tEnd@R,8.88
TEEK,0.87	tek,0.87	TICK,7.25	tIk,7.25
TEER,3.34	ter,3.34	TAR,3.14	t#R,3.14
TEMPEL,14.66	tEmp@l,14.66	TEMPLE,17.55	tEmpP,17.55
TERMIJN,6.29	tErmKn,6.29	TERM,17.45	t3m,17.45
TERREUR,3.11	tEr|r,3.11	TERROR,9	tEr@R,9
THEE,58.79	te,58.79	TEA,58.63	ti,58.63
TIENER,8.42	tin@r,8.42	TEENAGER,6.88	tin1_@R,6.88
TIJDPERK,8.67	tKtpErk,8.67	ERA,5.71	7r@,5.71
TIJGER,11.69	tKG@r,11.69	TIGER,18.53	t2g@R,18.53
TIK,6.81	tIk,6.81	TAP,14.75	t{p,14.75
TITEL,18.23	tit@l,18.23	TITLE,18.57	t2tP,18.57
TOEKOMST,103.45	tukOmst,103.45	FUTURE,103.49	fjuJ@R,103.49
TOERIST,3.91	turIst,3.91	TOURIST,4.65	t9rIst,4.65
TOGA,1.46	toGa,1.46	GOWN,6.55	g6n,6.55
TOMAAT,2.97	tomat,2.97	TOMATO,5.9	t@m#t5,5.9

TONG,31.9	tON,31.9	TONGUE,31.16	tVN,31.16
TOOST,5.79	tost,5.79	TOAST,33.47	t5st,33.47
TOREN,19.21	tor@,19.21	TOWER,22.84	t6@R,22.84
TOUW,26.25	tMw,26.25	ROPE,22.71	r5p,22.71
TRAAG,9.79	trax,9.79	SLOW,76.02	sl5,76.02
TRAAN,2.63	tran,2.63	TEAR,27	t7R,27
TRANT,0.53	trAnt,0.53	MANNER,11.53	m{n@R,11.53
TRAP,52.28	trAp,52.28	STAIRS,23.76	st8z,23.76
TRIEST,17.66	trist,17.66	SAD,63.37	s{d,63.37
TRIOMF,3.45	trijOmf,3.45	TRIUMPH,4.65	tr2@mf,4.65
TROFEE,4.37	trofe,4.37	TROPHY,7.55	tr5fI,7.55
TROMMEL,1.85	trOm@l,1.85	DRUM,8.47	drVm,8.47
TROPISCH,1.1	tropis,1.1	TROPICAL,3.2	trQpIkP,3.2
TROTS,111.46	trOts,111.46	PROUD,83.63	pr6d,83.63
TROTTOIR,1.28	trOtwar,1.28	SIDEWALK,5.47	s2dw$k,5.47
TROUW,52.34	trMw,52.34	LOYALTY,11.67	l4@ltI,11.67
TRUC,19.62	tryk,19.62	TRICK,47.27	trIk,47.27
TRUI,11.62	trL,11.62	SWEATER,13.8	swEt@R,13.8
TUIN,36.66	tLn,36.66	GARDEN,26.55	g#dH,26.55
TULP,0.48	t}lp,0.48	TULIP,0.78	tjulIp,0.78
TWAALF,46.49	twalf,46.49	TWELVE,18.82	twElv,18.82
TWEE,1007.83	twe,1007.83	TWO,1066.35	tu,1066.35
TYFOON,0.41	tifon,0.41	TYPHOON,1.47	t2fun,1.47
UITEEN,6.4	Lten,6.4	APART,47.02	@p#t,47.02
UITGANG,14.68	LtxAN,14.68	ENDING,15.92	EndIN,15.92
UNIE,2.08	yni,2.08	UNION,21.78	junj@n,21.78
UUR,348.69	yr,348.69	HOUR,162.29	6@R,162.29
VAAG,8.99	vax,8.99	VAGUE,4.53	v1g,4.53
VAANDEL,0.91	vand@l,0.91	BANNER,5.92	b{n@R,5.92
VAART,9.95	vart,9.95	VOYAGE,6	v4I_,6
VAARTUIG,1.56	vartLx,1.56	VESSEL,9.35	vEsP,9.35
VAATJE,0.59	vatj@,0.59	KEG,3.43	kEg,3.43
VADER,795.62	vad@r,795.62	FATHER,554.49	f#D@R,554.49
VANDAAG,369.61	vAndax,369.61	TODAY,433.8	t@d1,433.8
VARKEN,24.74	vArk@,24.74	PIG,39.14	pIg,39.14
VAT,19.05	vAt,19.05	BARREL,10.63	b{r@l,10.63
VEE,13.63	ve,13.63	CATTLE,13.22	k{tP,13.22
VEER,3.43	ver,3.43	QUILL,2.73	kwIl,2.73
VELD,15.96	vElt,15.96	FIELD,70.2	fild,70.2

VERDERF,2.38	v@rdErf,2.38	RUIN,28.53	rUIn,28.53
VERDRIET,26.48	v@rdrit,26.48	GRIEF,10.82	grif,10.82
VERF,14.48	vErf,14.48	PAINT,36.8	p1nt,36.8
VERHAAL,202.98	v@rhal,202.98	STORY,220.78	st$rI,220.78
VERKEER,15.28	v@rker,15.28	TRAFFIC,28.51	tr{fIk,28.51
VERKEERD,115.51	v@rkert,115.51	WRONG,523.1	rQN,523.1
VERKOOP,37.96	vErkop,37.96	SALES,12.71	s1lz,12.71
VERLIES,49.07	v@rlis,49.07	LOSS,29.12	lQs,29.12
VERSLAG,24.15	v@rslAx,24.15	REPORT,108	rIp$t,108
VET,18.52	vEt,18.52	GREASE,6.94	gris,6.94
VETE,1.12	vet@,1.12	FEUD,1.29	fjud,1.29
VETER,1.33	vet@r,1.33	SHOELACE,0.75	Sul1s,0.75
VIES,19.8	vis,19.8	DIRTY,24.56	d3tI,24.56
VIJF,281.57	vKf,281.57	FIVE,285.45	f2v,285.45
VIJVER,4.32	vKv@r,4.32	POND,6.33	pQnd,6.33
VINGER,28.91	vIN@r,28.91	FINGER,36.67	fINg@R,36.67
VIOOL,4.3	vijol,4.3	VIOLIN,4.75	v2@lIn,4.75
VIS,50.08	vIs,50.08	FISH,83.49	fIS,83.49
VISIE,8.23	vizi,8.23	VISION,23.88	vIZH,23.88
VLAG,17.79	vlAx,17.79	FLAG,17.49	fl{g,17.49
VLAK,47.73	vlAk,47.73	PLANE,95.53	pl1n,95.53
VLAM,7.94	vlAm,7.94	FLAME,9.04	fl1m,9.04
VLEES,61.67	vles,61.67	MEAT,43.65	mit,43.65
VLIEGER,3.8	vliG@r,3.8	KITE,2.29	k2t,2.29
VLOED,2.2	vlut,2.2	FLOOD,5.71	flVd,5.71
VLOEK,18.77	vluk,18.77	CURSE,18.22	k3s,18.22
VLUCHTIG,1.58	vl}xt@x,1.58	FUGITIVE,5.18	fju_@tIv,5.18
VOCHT,16.19	vOxt,16.19	LIQUID,7.75	lIkwId,7.75
VOD,1.51	vOt,1.51	RAG,4.78	r{g,4.78
VOERTUIG,12.1	vurtLx,12.1	VEHICLE,22.61	v7kP,22.61
VOET,50.81	vut,50.81	FOOT,64.92	fUt,64.92
VOLK,71.55	vOlk,71.55	PEOPLE,1102.98	pipP,1102.98
VONNIS,7.48	vOn@s,7.48	VERDICT,10.94	v3dIkt,10.94
VOORJAAR,2.38	vorjar,2.38	SPRING,31.31	sprIN,31.31
VORM,39.47	vOrm,39.47	SHAPE,30.24	S1p,30.24
VOS,7.59	vOs,7.59	FOX,21.61	fQks,21.61
VRAAG,436.76	vrax,436.76	QUESTION,198.35	kwEsJ@n,198.35
VREDE,65.01	vred@,65.01	PEACE,69.61	pis,69.61
VREEMD,126.67	vremt,126.67	STRANGE,86.43	str1n_,86.43

VRIEND,491.43	vrint,491.43	FRIEND,419.29	frEnd,419.29
VROUW,821.67	vrMw,821.67	WOMAN,434.63	wUm@n,434.63
VUIL,24.06	vLl,24.06	FOUL,14.47	f6l,14.47
VUILNIS,8.94	vLlnIs,8.94	GARBAGE,26.1	g#bI_,26.1
VUIST,6.77	vLst,6.77	FIST,7.35	fIst,7.35
WAARDE,25.63	ward@,25.63	WORTH,109.2	w3T,109.2
WAARHEID,189.39	warhKt,189.39	TRUTH,192.18	truT,192.18
WAGEN,77.82	waG@,77.82	WAGON,17.76	w{g@n,17.76
WALGING,1.12	wAlGIN,1.12	DISGUST,2.76	dIsgVst,2.76
WALNOOT,0.71	wAlnot,0.71	WALNUT,1.96	w$lnVt,1.96
WANG,7.89	wAN,7.89	CHEEK,7.16	Jik,7.16
WAPEN,140.59	wap@,140.59	WEAPON,46.65	wEp@n,46.65
WARMTE,10.5	wArmt@,10.5	WARMTH,4.45	w$mT,4.45
WASBEER,1.58	wAzber,1.58	RACCOON,1.43	r@kun,1.43
WEIDE,1.26	wKd@,1.26	PASTURE,1.53	p#sJ@R,1.53
WEINIG,110.45	wKn@x,110.45	LITTLE,1446.39	lItP,1446.39
WEKEN,148.3	wek@,148.3	SOAK,3.45	s5k,3.45
WELZIJN,5.83	wElzKn,5.83	WELFARE,7.88	wElf8R,7.88
WENK,0.39	wENk,0.39	HINT,9.2	hInt,9.2
WENS,59.39	wEns,59.39	WISH,235.12	wIS,235.12
WERELD,394.01	wer@lt,394.01	WORLD,455.22	w3ld,455.22
WERK,680.23	wErk,680.23	WORK,798.02	w3k,798.02
WERPER,1.42	wErp@r,1.42	PITCHER,3.24	pIJ@R,3.24
WET,80.45	wEt,80.45	LAW,116.31	l$,116.31
WETENSCHAP,23.3	wet@sxAp,23.3	SCIENCE,37.25	s2@ns,37.25
WEZEL,2.06	wez@l,2.06	WEASEL,4.9	wizP,4.9
WIEG,4.12	wix,4.12	CRADLE,2.84	kr1dP,2.84
WIEL,7.04	wil,7.04	WHEEL,27.06	wil,27.06
WIJN,60.44	wKn,60.44	WINE,60.35	w2n,60.35
WIJNSTOK,0.55	wKnstOk,0.55	VINE,2.1	v2n,2.1
WIJSHEID,11.09	wKshKt,11.09	WISDOM,11.08	wIzd@m,11.08
WIJZE,25.18	wKz@,25.18	SAGE,1.75	s1_,1.75
WINNAAR,23.6	wInar,23.6	WINNER,31.22	wIn@R,31.22
WISSEL,2.74	wIs@l,2.74	SWITCH,28.12	swIJ,28.12
WIT,33.48	wIt,33.48	WHITE,171.45	w2t,171.45
WOESTIJN,27.74	wustKn,27.74	DESERT,27.98	dEz@t,27.98
WOL,3.75	wOl,3.75	WOOL,3.16	wUl,3.16
WOLK,5.44	wOlk,5.44	CLOUD,11.75	kl6d,11.75
WOORD,129.09	wort,129.09	WORD,235.55	w3d,235.55

WORST,8.78	wOrst,8.78	SAUSAGE,7.78	sQsI_,7.78
WORTEL,6.11	wOrt@l,6.11	ROOT,10.47	rut,10.47
WRAAK,44.02	wrak,44.02	REVENGE,19.04	rIvEn_,19.04
ZAAL,15.41	zal,15.41	HALL,51.94	h$l,51.94
ZACHT,22.85	zAxt,22.85	SOFT,32.02	sQft,32.02
ZAK,96.87	zAk,96.87	POCKET,35.71	pQkIt,35.71
ZALF,1.69	zAlf,1.69	OINTMENT,1.63	4ntm@nt,1.63
ZALIG,9.15	zal@x,9.15	GORGEOUS,24.06	g$_@s,24.06
ZAND,19.99	zAnt,19.99	SAND,20.29	s{nd,20.29
ZEDIG,0.23	zed@x,0.23	MODEST,5.88	mQdIst,5.88
ZEE,67.8	ze,67.8	SEA,59.84	si,59.84
ZELDZAAM,10.02	zEltsam,10.02	RARE,21.31	r8R,21.31
ZETEL,3.06	zet@l,3.06	SEAT,78.78	sit,78.78
ZEVEN,103.52	zev@,103.52	SEVEN,104.45	sEvH,104.45
ZIEKTE,37.14	zikt@,37.14	SICKNESS,7.94	sIknIs,7.94
ZIEL,82.67	zil,82.67	SOUL,76.96	s5l,76.96
ZIJDE,18.8	zKd@,18.8	SILK,9.78	sIlk,9.78
ZILVER,9.95	zIlv@r,9.95	SILVER,31.75	sIlv@R,31.75
ZIN,172.65	zIn,172.65	SENTENCE,20.53	sEnt@ns,20.53
ZINLOOS,13.74	zInlos,13.74	USELESS,19.94	juslIs,19.94
ZITTEND,2.15	zIt@nt,2.15	SITTING,94.39	sItIN,94.39
ZOET,10.98	zut,10.98	SWEET,145.2	swit,145.2
ZOLDER,8.19	zOld@r,8.19	ATTIC,7.84	{tIk,7.84
ZOMER,42.9	zom@r,42.9	SUMMER,78.67	sVm@R,78.67
ZON,68.81	zOn,68.81	SUN,69.67	sVn,69.67
ZOOM,4.18	zom,4.18	HEM,0.82	hEm,0.82
ZOON,348.19	zon,348.19	SON,410.76	sVn,410.76
ZORG,218.82	zOrx,218.82	CARE,485.25	k8R,485.25
ZUID,5.12	zLt,5.12	SOUTH,64.47	s6T,64.47
ZUIVEL,0.64	zLv@l,0.64	DAIRY,2.78	d8rI,2.78
ZUIVER,10.63	zLv@r,10.63	PURE,24.92	pj9R,24.92
ZUS,131.03	z}s,131.03	SISTER,180.53	sIst@R,180.53
ZUSTER,71.51	z}st@r,71.51	NURSE,44.98	n3s,44.98
ZUURSTOF,13.13	zyrstOf,13.13	OXYGEN,13.88	QksI_@n,13.88
ZWAAN,1.99	zwan,1.99	SWAN,6.82	swQn,6.82
ZWAARTE,0.34	zwart@,0.34	GRAVITY,6.88	gr{v@tI,6.88
ZWAKTE,5.9	zwAkt@,5.9	WEAKNESS,8.9	wiknIs,8.9
ZWALUW,0.85	zwalyw,0.85	SWALLOW,12.73	swQl5,12.73
ZWART,57.31	zwArt,57.31	BLACK,167.94	bl{k,167.94

ZWEER,79.42	zwer,79.42	ULCER,2.57	Vls@R,2.57
ZWEET,12.87	zwet,12.87	SWEAT,21.86	swEt,21.86

14.2 Appendix B: Input words monolingual word recognition

Word	Frequency	Word	Frequency	Word	Frequency
ACID	9,96	FULL	166,9	PRAY	36,22
AGED	3,59	FURY	3,82	PREY	5,51
AGES	11,12	FUSE	4,86	PROP	3,69
ALLY	32,67	FUSS	7,14	PULL	146,45
ALSO	206,43	GAIN	13,73	PURE	24,92
ARCH	3,69	GALE	4,9	PUSH	70,55
AREA	74,92	GAME	233,84	QUIT	90,1
ARMY	85,69	GANG	30,14	RAFT	4,71
ASKS	16,71	GATE	32,04	RAGE	11,31
AUNT	55,2	GEAR	16	RAID	8,55
AWAY	730,9	GIRL	557,12	RAIL	4,57
BABE	35,57	GLAD	171,37	RAIN	48,9
BACK	2009,16	GLUE	5,88	RAKE	2,98
BAIL	17,35	GONE	296,76	RAMP	2,88
BAIT	9,73	GOWN	6,55	RAPE	16
BALD	9,73	GRAB	70,86	RARE	21,31
BANG	19,98	GRAY	21,12	RATE	24,92
BARE	8,33	GREW	29,27	READ	241,22
BARK	5,49	GREY	9,2	REAL	442,8
BARN	13,59	GRIM	4,78	REEL	3,25
BEAM	8,73	GROW	59,49	RELY	7,18
BEAT	131,69	GRUB	2,73	RENT	34,55
BEAU	5,84	HACK	8,55	RICE	15,08
BEEF	19,71	HANK	25,31	RIOT	6,49
BEEN	1736,73	HARM	31,78	RISE	27,43
BEEP	6,51	HAUL	7,08	ROAD	111,94
BELL	39,33	HAVE	6161,41	ROAM	3,39
BELT	24,35	HEAD	371,51	ROAR	4,02
BEND	15,06	HEAL	11,33	ROBE	8,49
BENT	7,22	HEAR	555,35	RODE	10,39
BILL	118,45	HEAT	40,08	ROOF	35,65
BIRD	45,45	HEIR	5,22	ROOM	439,51
BITE	40,78	HELM	5,35	ROOT	10,47
BLOW	97,57	HERB	4,98	ROPE	22,71
BODY	195,53	HERD	7,06	ROSY	3,31
BOIL	5,94	HIDE	69,69	RUDE	22,06
BOLD	7,55	HIGH	195	RUIN	28,53

BOND	31,27	HILL	37,55	RULE	48,14
BOOM	21,8	HINT	9,2	RUNS	32,1
BOOT	11,14	HOLD	436,73	RUSH	31,41
BORE	7,75	HOLE	58,22	SAFE	143,2
BOTH	295,33	HOLY	68,14	SAID	1108,45
BOUT	19,1	HOME	774,33	SALE	25,39
BOWL	21,45	HOOD	15,39	SALT	19,51
BRAG	3,51	HOOT	2,61	SAME	417,18
BUCK	33,75	HOSE	8,06	SANE	5,51
BULB	3,92	HOST	15,02	SANG	8,22
BULL	27,51	HUGE	48,37	SAVE	162,31
BUNK	6,27	HULK	4,08	SCAR	8,47
BURN	55,22	HULL	4,22	SEAL	14,75
BURY	20,67	HUMP	4,41	SEAT	78,78
BUSH	14,12	HUNG	28,16	SEED	7,57
BUSY	106,53	HUNK	5,16	SEEK	18,31
BUTT	38,57	HUNT	25,86	SEEM	139,82
BUZZ	15,96	HURT	246,35	SEES	37,24
CAGE	20,27	HUSH	7,61	SELL	92,25
CALL	861,39	IDLE	2,76	SEND	179,78
CAME	463,73	INCH	12,37	SENT	138
CANE	8,33	IRON	17,94	SHED	10,98
CAPE	8,24	ITCH	4,18	SHIN	3,08
CARD	85,43	JACK	251,59	SHOP	53,55
CARE	485,25	JAIL	70,63	SHOW	488,35
CASE	282,41	JERK	33,14	SHUT	263,82
CASH	72,43	JOIN	83,43	SIDE	200,92
CAST	23,14	JOKE	73,02	SIGH	3,39
CAVE	13,98	JUMP	69,82	SIGN	133,27
CHAP	6,35	JUNK	15,37	SILK	9,78
CHAT	16,27	JUST	4749,14	SINK	16,92
CHEW	9,06	KEEN	4,47	SITE	19,22
CHIN	12,69	KEEP	702,86	SIZE	46,14
CHOP	13,61	KEPT	89,39	SKIN	44,04
CHOW	6,86	KICK	73,41	SKIP	21,1
CITY	169,1	KIDS	301,1	SLAM	5,8
CLAN	4,1	KILL	452,57	SLIM	11,86
CLAY	12	KIND	590,69	SLIP	25,88
CLIP	5,69	KNEW	368,96	SLOT	5,49
COAL	6,57	KNOW	5721,18	SLOW	76,02
COAT	42,08	LACE	3,71	SLUG	4,96

COCK	11,37	LACK	17,75	SNAP	17,39
COIN	9,75	LADY	217,08	SOAK	3,45
COLD	130,16	LAKE	36	SOAP	15,2
COMB	6,06	LAME	10,92	SODA	19,84
CONE	2,92	LANE	33,41	SOFA	5,86
COOK	45,57	LASS	2,67	SOFT	32,02
COOL	195,88	LAWN	12,35	SOIL	7,78
COOP	10,35	LAWS	17,02	SOLD	52,06
COPE	3,25	LAZY	11,59	SOLE	5,25
COPY	52,27	LEAF	5,2	SONG	93,69
CORE	9,82	LEAN	10,37	SOON	257,65
CORN	14,22	LEAP	6,67	SORE	15
COUP	2,61	LEFT	484,45	SOUL	76,96
COVE	3,25	LENT	3,73	SPOT	61,57
CRAP	61,78	LESS	111,1	STAB	8,76
CREW	47,53	LIAR	35,14	STAY	515,65
CRIB	6,2	LIED	43,8	STIR	5,9
CROP	4,86	LIFE	796,65	STUD	6,94
CROW	4,45	LIKE	3998,96	SUCH	291,22
CULT	4,45	LIMB	4,67	SUCK	34,88
CURB	4,1	LIMP	3,67	SUIT	68,61
CURE	20,84	LION	15,35	SUNK	4,65
CUTE	87,75	LOAD	29,22	SURE	1099,82
DAME	13,76	LOAF	4,47	SWAP	3,63
DAMN	283,53	LOAN	19,86	SWIM	31,8
DAMP	2,92	LOCK	56,57	TAIL	23,9
DARK	88,61	LOFT	3,65	TAKE	1891,04
DARN	13,9	LONE	5,25	TALE	12
DASH	6,04	LOOK	1947,27	TALK	855
DATA	25,61	LOOP	6,76	TALL	32,33
DAWN	25,51	LOOT	3,55	TAPE	68,84
DAYS	305,73	LORD	138,16	TEAR	27
DEAL	261,37	LOSE	164,35	TELL	1724,49
DEAR	223,43	LOSS	29,12	TEND	12,27
DEBT	14,22	LOST	274	TEST	84,08
DEER	8,71	LOVE	1114,98	THEN	1489,53
DEFY	3,1	LUCK	153,73	THEY	4102,94
DENY	21,39	LUMP	3,55	THIN	20,18
DESK	43,9	LURE	4,25	THIS	7978,73
DICE	10,45	MADE	561,29	TIED	26,75
DIED	157,22	MAID	22,82	TILL	166,73

DIME	12,06	MAIN	42,73	TINY	32,22
DIRT	25,69	MALE	33,94	TIRE	12,37
DISH	11,45	MALL	18,9	TOLD	699,59
DISK	6,63	MANY	359,43	TOMB	5,63
DOLL	24,76	MARK	82,02	TOOK	342,24
DONE	485,04	MATE	29,24	TOOL	10,75
DOOM	5,61	MATH	16,39	TORE	8,29
DOPE	16,08	MEAL	28,86	TORN	10,84
DOVE	5,57	MEAN	1243,98	TOWN	247,92
DOWN	1490,27	MEAT	43,65	TRAP	23,84
DRAG	26,45	MEET	352,27	TRAY	8,04
DRAW	40,41	MERE	7,86	TREE	65
DRUM	8,47	MESS	78,14	TREK	4,49
DUCK	24,76	MILL	9,53	TRIP	82,39
DUKE	25,04	MIND	484,61	TRUE	253,35
DULL	12,08	MINK	3,71	TUBE	16,43
DUMB	46,96	MOCK	5,37	TUCK	7,96
DUMP	28,82	MOLD	4,29	TUNE	15,61
DUST	23,84	MOOD	34,04	TURF	4,27
DUTY	50,96	MORE	1298,59	TURN	306,47
EACH	253,25	MOVE	418,14	TWIN	10,43
EARL	20,43	MUCH	973,25	UGLY	42,16
EARN	15,35	NAVY	25,69	UNIT	36,18
EASE	19,1	NEAT	12,45	UNTO	7,63
EASY	265,71	NEED	1294,9	UPON	62,73
EDDY	3,61	NEXT	452,75	URGE	6,65
EDGE	23,51	NICE	649,51	USED	344,14
ELSE	449,16	NINE	67,47	VAIN	6,51
ENVY	9,55	NONE	110,61	VAST	6,1
EVEN	875,92	NOON	18,12	VEIL	2,96
EVER	709,22	NOSE	69,75	VEIN	3,59
EVIL	73,16	NUDE	5,9	VENT	4,41
EXIT	15,57	NUMB	4,88	VERY	1241,25
FACE	289,16	NUTS	53,51	VICE	18,63
FADE	5,61	OATH	9,88	VIEW	38,53
FAIL	24,59	OBEY	8,94	VILE	4,43
FAIR	94,75	ONCE	344,88	VOID	4,1
FAKE	36,33	ONLY	1083,71	VOTE	34,33
FAME	8,65	ONTO	36,69	WAGE	3,12
FARM	30,04	ORAL	4,31	WALK	215,86
FAST	137,45	PACE	9,57	WALL	70,69

FATE	26,96	PACT	3,76	WANT	2759,18
FEAR	69,08	PAGE	37,49	WARD	14,25
FEED	42,39	PAID	85,67	WARN	25,35
FEEL	627,24	PAIN	97,94	WARP	5,49
FEET	120,73	PALE	8,02	WAVE	21,25
FELL	73	PART	261,51	WEAR	109,33
FELT	119,82	PAST	123,76	WEEP	5,49
FILL	43,94	PEAK	5,94	WELL	2990,65
FINK	4,04	PECK	3,53	WENT	411,51
FIRE	215,49	PEEL	5,35	WHEN	2034,1
FIVE	285,45	PEEP	4,43	WHIP	13,16
FLAW	3,04	PEST	2,86	WHIZ	3,9
FLEA	3,31	PICK	198,39	WHOM	35,2
FLED	4,39	PILE	13,18	WIDE	23,8
FLEW	15,39	PIMP	8,63	WIFE	348,92
FLIP	14,27	PINE	6,2	WILL	2123,65
FLOW	13,75	PINK	28,47	WING	20,24
FOAM	3,51	PITY	23,51	WIPE	16,88
FOLD	8,63	PLAY	354,53	WIRE	27,51
FOLK	7,16	PLEA	6,84	WISH	235,12
FOOD	154,43	POEM	13,65	WITH	5048,33
FOOL	89,33	POET	9,22	WOOD	27
FORE	2,65	POND	6,33	WORE	21,2
FORM	42,75	POOL	46,98	WORN	10,08
FOUL	14,47	POOR	129,08	WRAP	17,8
FOUR	255,78	POPE	10,71	YANK	4,71
FREE	177,53	PORK	10,53	YARD	25,06
FROM	2039,06	PORT	14,53	YELL	18,41
FUEL	17,18	POUR	15,12	ZERO	21,45

14.3 Appendix C: Input words bilingual word recognition

Word	Frequency	Word	Frequency	Word	Frequency
AARD	15,32	KAAR	0,21	REIS	90,37
ADEM	53,15	KAAS	22,85	REUS	7,45
ADER	3,27	KAFT	0,71	RIEM	14,16
AKTE	4,18	KALK	1,01	RIET	0,87
ALFA	3,73	KALM	69,43	RIJK	79,15
ALOM	1,39	KAMP	41,73	RIJM	0,73
ALZO	0,48	KANO	2,24	RIJP	8,1
AMBT	2,36	KANS	264,38	ROEM	7,98
ARME	89,02	KANT	312,67	ROEM	7,98
AULA	0,69	KAST	30,05	ROEP	48,53
BAAI	4,51	KEET	1,17	ROER	7,8
BAAS	167,21	KERK	79,49	ROMP	6,47
BAAT	2,47	KERN	9,26	ROND	172,65
BARS	2,7	KERS	1,62	ROOD	48,59
BAST	0,41	KEUR	2,4	ROOS	11,71
BEDE	0,3	KEUS	61,61	ROTS	14,66
BEEK	3,43	KIES	35,35	ROUW	3,96
BEHA	4,12	KIJK	1049,77	ROZE	19,09
BERG	34,3	KIST	27,67	RUIG	5,05
BERM	0,69	KLAS	48,98	RUIG	5,05
BEST	306,52	KLEM	10,47	RUIL	15,39
BIER	53,35	KLEP	5,15	RUIS	1,3
BIES	0,11	KLIK	3,04	RUND	3,54
BIJL	9,26	KLIM	10,79	SAAI	31,47
BITS	0,82	KLIP	0,64	SAUS	9,56
BLAD	11,14	KLOK	23,9	SEIN	3,11
BLIJ	277,3	KLUS	37,78	SEXE	0,21
BLOK	11,62	KNAL	14,82	SLAG	63,94
BODE	1,67	KNAP	68,05	SLAG	63,94
BOEF	6,2	KNIE	10,24	SLEE	4,18
BOEK	150,93	KNIK	1,81	SLET	28,58
BOEL	52,37	KNOP	16,46	SMAK	1,92
BOER	14,77	KOEK	3,27	SMAL	2,08
BONK	0,57	KOEL	8,76	SMAL	2,08
BOOG	8,55	KONT	73,57	SMID	2,58
BOOR	3,45	KOOI	13,81	SNIT	0,3
BORD	27,3	KOOK	7,52	SNOR	9,95
BORG	11,23	KOOL	4,73	SOEP	17,84
BOUW	12,33	KOOP	48,34	SOMS	349,86

BOZE	14,89	KOOR	6,88	SPEK	5,35
BRES	0,69	KORT	47,52	SPEL	95,11
BRIL	24,49	KOST	80,4	SPUL	68,01
BROK	1,65	KOUD	95,95	STAD	272,61
BRON	29,93	KRAB	3,86	STAF	11,62
BRUG	44,07	KRAS	2,15	STAL	22,07
BUIK	26,09	KRAT	3,82	STAM	12,46
BUIL	0,91	KROM	2,84	STEK	1,3
BUIT	10,4	KROP	2,84	STEL	214,94
BULT	3,41	KRUK	2,56	STER	43,77
BUUR	3,06	KUIL	3	STIL	160,81
COLA	11,46	KUIP	0,85	STIP	1,56
DAAD	16,65	KUIT	0,5	STOF	29,25
DANS	37,82	KUST	31,88	STOK	14,48
DEEG	1,85	KUUR	0,85	STOM	103,91
DEUK	2,65	KWIK	1,67	STRO	1,78
DEUN	0,07	LAAG	31,1	STUK	175,67
DEUR	247,48	LAAN	1,23	TAAI	9,03
DIEF	29,96	LAAT	2032,73	TAAK	42,63
DIEP	68,4	LACH	46,72	TAAL	36,29
DIER	28,1	LADE	2,26	TAND	8,07
DING	282,24	LANG	596,85	TAST	5,4
DODE	59,85	LANS	1,9	TEEN	7,39
DOEK	8,83	LEED	12,21	TEER	3,34
DOEL	77,45	LEEG	61,74	TEIL	0,14
DOEN	2677,85	LEEM	0,27	TEUG	0,87
DOES	0,07	LEEP	0,05	THEE	58,79
DOOD	992,15	LEUK	595,57	TIEN	207
DORP	53,99	LIED	21,27	TIET	2,33
DREK	0,62	LIEF	113,95	TIET	2,33
DRIE	526,78	LIGA	0,57	TIJD	898,46
DRUK	207,11	LIJF	23,65	TIJD	898,46
DUIF	5,35	LIJK	74,46	TOCH	1693,53
DUIK	8,69	LIJM	3,98	TOEN	1086,55
DUIM	10,79	LIJN	68,72	TOGA	1,46
DUNK	2,38	LINT	7,98	TONG	31,9
DURF	47,41	LOEP	0,21	TOON	34,67
DUUR	34,53	LOGE	1,74	TOUW	26,25
EDEL	1,14	LOOD	5,03	TRED	0,57
EEND	8,99	LUID	9,08	TROM	0,5
EENS	1392,11	LUIK	6,68	TROS	0,34

EEUW	25,15	LUXE	10,11	TRUC	19,62
EIND	83,17	MAAG	23,55	TRUI	11,62
ENIG	51,75	MAAK	648,19	TRUT	40,11
ERAF	63,44	MAAL	17,29	TUIG	15,64
EREN	7,57	MAAN	42,1	TUIN	36,66
ERIN	7,57	MAAR	8385,73	TWEE	1007,51
EROP	158,18	MAAT	69,18	UNIE	2,08
ERWT	0,55	MAIS	5,55	VAAK	180,02
ETEN	445,92	MALS	1,12	VAAR	2,54
ETUI	0,07	MAND	4,3	VAAS	4,57
EZEL	11,41	MANK	2,58	VAAT	1,42
FASE	15,55	MARS	20,67	VALK	2,58
FEIT	43,04	MEEL	1,69	VEEG	6,7
FERM	0,18	MEER	2035,88	VELD	15,96
FLES	45,71	MEID	114,52	VERF	14,48
FOOI	10,93	MELK	39,7	VERS	17,86
FORD	7,36	MENS	144,66	VERS	17,86
FOTO	119,16	MERG	0,69	VIER	242,08
FOUT	165,27	MERK	19,12	VIES	19,8
FUIK	0,16	MIER	2,54	VIJF	281,57
FUST	0,11	MIJL	15,32	VLAG	17,79
GAAF	39,22	MIJN	4412,02	VLAK	47,73
GAAI	0,32	MOED	41,05	VLAK	47,73
GAAN	2650,41	MOOI	616,06	VLAM	7,94
GAAR	2,77	MOUW	4,18	VLEK	6,77
GADE	0,55	MUIL	1,88	VLOT	6,43
GANS	5,88	MUIS	11,14	VLUG	48,05
GAST	56,26	MUNT	10,89	VOEG	5,03
GAUW	60,55	MURW	0,48	VOET	50,81
GEEL	9,56	MUTS	4,46	VOLK	71,55
GEIL	11,8	NAAM	470,6	VONK	2,22
GEIT	8,07	NAAR	4447,55	VORK	5,19
GELD	793,61	NAZI	5,69	VORM	39,47
GESP	0,96	NEEF	48,48	VOUW	1,72
GIDS	11,39	NEEN	20,76	VRIJ	232,91
GLAS	57,15	NEUS	70,04	VUIL	24,06
GOED	3488,11	NIER	3,91	VUUR	100,57
GOOT	6,06	NIET	18323,98	WAAS	0,91
GOUD	61,99	NIJD	0,34	WALM	0,16
GRAF	34,53	NIKS	831,61	WALS	2,4
GRAP	51,02	NOOD	15,48	WANG	7,89

GRAS	18,75	NOTA	1,1	WARS	1,07
GROS	0,34	OBER	10,02	WEER	1136,95
GROT	17,45	OGEN	260,85	WEES	233,69
HAAI	9,44	OLIE	25,25	WENK	0,39
HAAK	12,65	OOIT	465,25	WENS	59,39
HAAL	302,29	OPOE	0,62	WERF	1,42
HAAN	4,28	ORDE	50,49	WERK	680,23
HAAR	2975,87	OUWE	75,26	WIEL	7,04
HAAS	2,17	PAAR	434,44	WIER	2,54
HAAT	151,32	PAKT	35,01	WIJF	21,79
HALS	7,5	PAND	5,85	WIJK	9,58
HEEN	396,09	PAUS	12,62	WIJN	60,44
HEER	104,14	PAUW	0,96	WIJS	36,25
HEIL	5,72	PEIL	1,72	WILG	0,14
HEKS	26,76	PELS	0,78	WOLK	5,44
HELD	59,64	PERS	37	WOND	14,02
HELM	11,09	PIEK	4,94	WORP	6,72
HEMD	11,98	PIJL	7,11	WOUD	5,19
HERT	6,13	PIJP	13,81	WRAK	9,33
HEUP	4,09	PION	3,18	ZAAD	6,33
HEUS	49,83	PLAK	3,13	ZAAG	3,54
HIEL	0,85	PLEK	177,41	ZAAK	239,34
HIER	4516,62	PLEK	177,41	ZAAL	15,41
HOED	35,95	PLUK	2,84	ZALM	5,28
HOEK	49,92	POEL	1,78	ZAND	19,99
HOER	55,32	POES	6,61	ZANG	1,65
HOMP	0,27	POLS	13,65	ZEEP	14,22
HOND	168,65	PONT	0,66	ZEER	149,85
HOOG	63,23	POOS	3,48	ZEER	149,85
HOOI	2,88	POOT	6,95	ZEGE	4,67
HOUT	23,58	POOT	6,95	ZEIL	6,91
HUID	39,56	PRAT	0,46	ZIEK	129,2
HULP	239,77	PRAT	0,46	ZIEL	82,67
HUUR	28,88	PRET	9,03	ZIEN	1538,8
IDEE	482,99	PRIK	4,6	ZIJN	7968,23
IETS	2245,19	PRIL	0,25	ZOEK	199,04
INKT	5,17	PROF	3,45	ZOEN	9,99
JAAR	762,67	PUIN	2,38	ZOET	10,98
JEUK	2,58	PUUR	23,35	ZOOL	0,91
JOCH	18,77	RAAD	83,01	ZOON	348,19
JONG	91,61	RAAF	1,74	ZORG	218,82

JOOD	14,34	RAAM	70,84	ZOUT	15,5
JULI	11,8	RAAR	84,89	ZUID	5,12
JUNI	12,94	RANG	9,19	ZUIL	0,55
JURK	55,75	RAUW	4,67	ZULK	7,02
KAAK	5,12	REDE	5,21	ZULT	219,33
KAAL	8,23	REEP	2,65	ZUUR	10,18
KAAP	1,1	REET	53,21	ZWAK	35,4

14.4 Appendix D: Input word for word translation from Pruijn (2015)

Dutch NC LF	Dutch C LF	Dutch C HF	Dutch NC HF
0:AREND	0:BAKKER	0:BIER	0:AUTO
0:BEKER	0:BALKON	0:BOEK	0:AVOND
0:BIJL	0:BIJBEL	0:DOCHTER	0:BEDRIJF
0:BOER	0:BOT	0:EINDE	0:BOOS
0:DUIF	0:BRUID	0:GOUD	0:BUREAU
0:GEEL	0:BRUIN	0:GROND	0:DAK
0:GEZOND	0:DIEF	0:HOOFD	0:DONKER
0:HAAI	0:DOMEIN	0:HUIS	0:DORP
0:HERFST	0:DROOG	0:KAT	0:GEBOUW
0:IJZER	0:GEIT	0:KOFFIE	0:GELD
0:JAGER	0:HONING	0:MAAN	0:GEVAAR
0:KNOP	0:KLIMAAT	0:MOEDER	0:GEVOEL
0:KOORTS	0:MUIS	0:MOND	0:GEZICHT
0:LAWAAI	0:NATUUR	0:MUZIEK	0:HOND
0:LEPEL	0:OOR	0:NAAM	0:JONGEN
0:LIED	0:PAREL	0:NACHT	0:KANTOOR
0:MAAG	0:PEPER	0:PIJN	0:LICHAAM
0:MEISJE	0:PIJP	0:PRIJS	0:MES
0:MIER	0:PIRAAT	0:RIJK	0:MOE
0:MOUW	0:RAUW	0:RIVIER	0:MUUR
0:NIER	0:REGEN	0:STRAAT	0:OORLOG
0:PERZIK	0:SAUS	0:TAFEL	0:PAARD
0:PITTIG	0:SOK	0:THEE	0:RUIMTE
0:SLAGER	0:SOM	0:VADER	0:SCHOON
0:SPIER	0:TROFEE	0:VOET	0:SLEUTEL
0:STIER	0:VELD	0:VRIEND	0:STAD
0:VERF	0:VLAM	0:WERELD	0:TANTE
0:VERKEER	0:VOS	0:WERK	0:VERHAAL
0:VIES	0:WARMTE	0:WIJN	0:VRAAG
0:VLOEK	0:WEZEL	0:WOORD	0:VROUW
0:WOLK	0:ZILVER	0:ZEE	0:ZORG
0:ZACHT	0:ZWEET	0:ZOON	0:ZWART

English NC LF	English C LF	English C HF	English NC HF
0:ANT	0:BAKER	0:BOAT	0:ANGRY
0:AUTUMN	0:BIBLE	0:BOOK	0:ART
0:AXE	0:BRIDE	0:BOSS	0:BLACK
0:BULL	0:BROWN	0:COFFEE	0:BODY
0:CLOUD	0:CORAL	0:DAUGHTER	0:BOY
0:COWARD	0:CULTURE	0:FATHER	0:BUILDING
0:CURSE	0:CURL	0:FIELD	0:CAR
0:EAGLE	0:DOMAIN	0:FIGURE	0:CLEAN
0:FARMER	0:EAR	0:FOOT	0:DARK
0:FEVER	0:FLAME	0:FRIEND	0:DIRTY
0:GARDEN	0:FOX	0:GROUND	0:DOG
0:GARLIC	0:METHOD	0:HAIR	0:EYE
0:GLOVE	0:MOTIVE	0:HONEY	0:FACE
0:HEALTHY	0:MOUSE	0:HOUSE	0:FEELING
0:HUNTER	0:PANIC	0:MIDDLE	0:GIRL
0:JUICE	0:PEARL	0:MINUTE	0:HORSE
0:KIDNEY	0:PEPPER	0:MOTHER	0:KEY
0:LACK	0:PIPE	0:MOUTH	0:LOW
0:MIRROR	0:PIRATE	0:MUSIC	0:MONEY
0:MUSCLE	0:RAW	0:NAME	0:MORNING
0:PEACH	0:RICE	0:NIGHT	0:OFFICE
0:PENCIL	0:SAUCE	0:PAIN	0:QUESTION
0:PLATE	0:SILVER	0:POLICE	0:SHOWER
0:RABBIT	0:SLAVE	0:RICH	0:SONG
0:SCAR	0:SWEAT	0:STREET	0:SPACE
0:SCIENCE	0:TASK	0:SUN	0:STORY
0:SHARK	0:THIEF	0:TABLE	0:TIRED
0:SLEEVE	0:TIGER	0:TEA	0:UNCLE
0:SOFT	0:TROPHY	0:WORD	0:VOICE
0:SPICY	0:VAGUE	0:WORK	0:WALL
0:SPOON	0:VISION	0:WORLD	0:WINDOW
0:YELLOW	0:WARMTH	0:YEAR	0:WOMAN