# Simulating the learning of languages: How bilinguals learn cognates, false friends, and words of different frequencies

Alex Bijsterveld

August 23, 2010

BA thesis Artificial Intelligence
Author: Alex Bijsterveld
Student number: 0520195
Supervisors: Ton Dijkstra & Ida Sprinkhuizen-Kuyper
Radboud University Nijmegen
the Netherlands
*alexbijsterveld@student.ru.nl*
August 23, 2010

**Abstract**

With a distributed connectionist model I simulated the acquisition of a second language by early and late bilinguals and explored the differences between these processes. I also added cognates, false friends, and translation equivalents to the lexicons to investigate if the model was able to mimic the reaction time effects for such words as found in empirical studies.

The model showed that the process of learning a second language was slowed down by the knowledge of a native language. It was also able to simulate a cognate facilitation effect, a cognate similarity effect, an interference effect for false friends, and a word-frequency effect. These conclusions contribute to the notion of the language-nonselective nature of the bilingual word recognition system.

# 1. Introduction

In the last half century, many studies have been done in the field of lexical processing. Initially, these studies investigated the performance of monolingual language users, but since the nineties, more and more studies have appeared that consider how language users process words in a non-native language. In particular, researchers wondered about two basic issues. First, do bilingual speakers have two separate lexicons for each of their languages or one large bilingual lexicon? Second, what are the underlying mechanisms that allow lexical access and lexical selection in bilinguals? A host of empirical studies have been done to investigate these issues (see Dijkstra, 2009, for a review).

More recently, computational techniques from Artificial Intelligence have been introduced into this area of research. Several computational models for bilingual word processing have already been developed, especially of a localist or distributed connectionist type. These have offered successful accounts with respect to the word retrieval mechanisms in bilinguals. In the present study, we will consider these computational models and present a series of simulations with a new distributed computer model for second-language acquisition. Before we discuss the presently available models, we will first consider some empirical studies making use of the special types of words that will later be stimuli in the simulations.

Dijkstra, Miwa, Brummelhuis, Sappelli, and Baayen (2010) consider one important type of special types of words in relation to bilingual processing. These are **cognates,** words that have (almost) the same form and meaning in the native language and the new language (e.g., *lamp-lamp*). These cognates have an advantage during processing, called the cognate facilitation effect. In fact, a cognate is a special type of translation equivalent that has a large or complete overlap in form across languages. They can be identical or non-identical (e.g., *lamp-lamp* vs. *tomaat-tomato*). In the lexical decision task, Dijkstra et al. found a similarity effect for cognates. More similar translation equivalents were recognized more quickly than translation equivalents with less or no similarity at all. In the language selection task, orthographic similarity had an inhibitory effect on reaction times to cognates. The existence of these effects suggest that translation equivalents do not have a totally independent representation in the two languages. Dijkstra et al. propose that cognates share their meaning, but have language-specific orthographic representations.

A different type of special words are **false friends.** These words are alike in their written form for different languages, but they have a different meaning (e.g., *ramp-ramp*, which means ‘disaster’ in Dutch). Dependent on the relative frequency of their readings in the two languages they suffer from an interference effect. This effect also contributes to the notion of the language-nonselective nature of the bilingual word recognition system. In the paper of Dijkstra, van Jaarsveld, and ten Brinke (1998) three lexical decision task experiments were reported involving false friends to obtain some information about language selective and non-selective access. In the first experiment, only English words and nonwords were included and the participants had to decide if a shown word was an English word. Here they did not find any reaction time effect for the false friends relative to control words. In the second experiment, the authors added Dutch words to the stimulus material and instructed the participants to treat the Dutch words as nonwords. The task was the same as in the first experiment. In this case, there was an inhibition effect for the false friends relative to matched one-language control words. In the last experiment, they used the same stimulus material, but the task was slightly different from the first two experiments. The participants had to give a “yes” response when the word presented was English or Dutch and a “no” response when the word did not exist in either languages. Now a facilitatory effect arose for the false friends. The last two experiments support the hypothesis of language nonselective access.

The size of the inhibition effects in the second experiment and the facilitation effects in the last experiment depended on the **frequency** of the false friends in the native language (Dutch) relative to the frequency in the second language (English). High-frequency readings become available more quickly than low-frequency readings and will therefore inhibit the lexical decision to English in the second experiment and facilitate language selection in the last experiment, where a response could be based on both English and Dutch.

In the last decade, a series of models for bilingual word processing has been developed in which the processing of these types of special words is accounted for. Some of these models have been implemented as localist-connectionist models or distributed-connectionist models. We will discuss these models in the next section.

*Localist connectionist models*

An example of a localist connectionist model is the Bilingual Interactive Activation (BIA) model (Dijkstra & van Heuven, 1998; van Heuven, Dijkstra, & Grainger, 1998). This model is successful in explaining and modeling lexical selection in bilinguals. It provides a precise simulation of interlingual orthographic neighbor word effects, between-language masked orthographic prime effects, and accounts for interlingual homograph recognition effects. However, the BIA model's use of language tags has been criticized. When compared to the BIA model, the localist Bilingual Model of Lexical Access (BIMOLA, Léwy & Grosjean, in press) shows that different theoretical assumptions might be needed to simulate empirical effects in the visual and auditory modalities.

Another localist model is the SOPHIA model (Semantic, Orthographic, and Phonological Interactive Activation model) (Van Heuven & Dijkstra, in preparation). This model is a further extension of the BIA model (Dijkstra & van Heuven, 2002). It can mimic some well-known effects in monolingual visual word recognition. These are priming effects, consistency effects between orthographic and phonological codes, pseudohomophone effects, and neighborhood effects. The main power of SOPHIA is to simulate the effects of neighbor words that share their orthographical rime with the target word.

Femke Haga (2010) wondered if learning a second language would affect the first learned language and vice versa. To investigate this issue, she performed simulations with the BIA model using four-letter words. She found that, depending on the stage of learning, performance in the second learned language indeed had its influence on the first learned language and vice versa. We will come back to this simulation work with a localist model later and compare it to our own simulations with a distributed model.

*Distributed-connectionist models*

Thomas (1997a, 1997b, 1998) examined how to extend the distributed models of monolingual processing to the bilingual domain. He explored the Single Network hypothesis, which holds that interference effects between languages are a consequence of the storage of two languages in only one representing source. He developed the Bilingual Single Network (BSN) model to show behavior that illustrated the language independence of lexical representations as well as interference effects. With respect to the independence effects, interlingual homographs showed intra-language frequency effects. There was an absence of between-language long term priming effects for words with the same meaning in both languages. With respect to interference effects, the model demonstrated disadvantages for interlingual homographs in comparison to cognate homographs. In the unbalanced network, a facilitatory effect was observed for cognate homographs in the second language (L2). The model was also able to imitate between-language semantic priming effects by using a shared semantic output layer. However, because the learning model received two artificial languages as input, language tags were used. These language tags allowed the model to distinguish the first (L1) and second (L2) language, while the model probably would not be able to do so without these language tags.

With a different distributed bilingual model, referred to as Bilingual SRN, French and Jacquet (2004) wanted to find out if word order information in sentences would provide enough information to distinguish two languages. They found that when language switch was included in sentences with a low probability (0.1%), the order of words alone was enough to create different representations of the two languages. Thus, the Bilingual SRN model supports the hypothesis that bilingual memory is organized in only one distributed lexicon.

Erik Lormans (2010) extended the BSN model of Thomas (eBSN) and wondered if he could obtain the same results by using words from natural languages. As languages he chose English and Dutch. He predicted differences in learning time for cognates, false friends, and translation equivalents as compared to control words. In the mixed simulations with two languages, Lormans showed a facilitation effect for cognates relative to control words, as well as an inhibitory effect for both false friends and translation equivalents. The presence of these effects is in line with empirical findings of language non-selective lexical access in bilinguals.

In the present research, I will also consider how the processing of languages in the case of bilinguals can be simulated by a distributed connectionist model. My model is based on the model of Erik Lormans, eBSN, the extended BSN model. Bilinguals can have a native language (L1) and acquired a second language (L2) later (late bilingual), or they can be raised with two languages from birth (early bilingual). I wish to explore any differences in simulations of the processes of acquisition of a (second) language in late bilinguals, early bilinguals, and monolinguals.

I will conduct these simulations with English words and Dutch words that have different frequencies of usage. In this way, I can also investigate the effect of the frequency of a word on the learning process of that word.

In the English and Dutch lexicons, I add some special types of words: false friends, cognates, and translation equivalents with varying orthographic overlap. Can I demonstrate a cognate facilitation effect in my simulations? Do the false friends in my simulations also experience an inhibitory effect on the amount of time needed to learn them? And does the amount of overlap in the translation equivalent group correlate with the amount of time needed to learn them? In sum, will I be able to mimic the reaction time effects found in empirical studies for these word types in simulations of how they are learned?

In the following sections, I will first explain the model and the used lexicons in chapter 2. In chapter 3 I will perform some basic simulations to show that my model is able to simulate learning processes of monolinguals and bilinguals at all. In chapter 4, I will make some comparisons between these different types of bilingual simulations and monolingual simulations. After that I will have a closer look at the learning processes for cognates, false friends, and translation equivalents in chapter 5. I will also take a look at the learning processes of the different orthographic overlap groups belonging to the translation equivalents. In chapter 6 I will consider the different frequency groups and in the end, I will discuss my results and compare them to earlier research.

# 2. Method

The model I will use for my research is based on the eBSN-model of Erik Lormans. It is an extended version of the BSN-model of Thomas. The model is a single network that is trained to generate meanings of the words in two languages and can be used to find evidence for between language similarity effects.

For my research I only used four-letter words. The lexicon of words with their corresponding frequency I used are originally from the CELEX database (Baayen, Piepenbrock, & Van Rijn, 1993). I translated the English words to Dutch and categorized them into cognates, false friends, translation equivalents, and control words.

Because I knew from each word in my lexicon how many times it occurred in a collection of articles and papers, I could create a frequency element in my lexicons. The frequency of a word ranged from only two times up to twenty thousand times. I decided to create three types of frequency groups: high frequency, middle frequency and low frequency. In the collection of words there were more lower than higher frequencies, so simply dividing twenty thousand in three equal parts would mean that most of the words would belong to the first group with frequencies lower than approximately seven thousand. To avoid this I twice took the logarithm of the frequencies. This resulted in a group of words with modified frequencies ranging from zero to seven. To roughly create three equally divided groups I found out that the best division would be to put words with modified frequencies zero and one in the low-frequency group, words with modified frequencies two and three in the middle-frequency group and words with modified frequencies four, five, six and seven in the high-frequency group. Then I create my lexicon such that the low-frequency words appear only once in the lexicon, the middle-frequency words appear twice in the lexicon and the high-frequency words each appear four times in the lexicon.

Each cognate, false friend and translation equivalent I matched with a control word. The control word belonged to the same frequency category and each letter from this control word was different from the letter at the corresponding position in the special word. For my simulations I used an English lexicon, a Dutch lexicon, and a Dutch-English lexicon. The Dutch and the English lexicon both consisted of 40 unique cognates, 40 unique false friends, 80 unique translation equivalents (20 for each orthographic overlap category) and at most 160 unique control words (128 for the English lexicon and 122 for the Dutch lexicon). The reason I use a maximum is because this number is decreased by the fact that sometimes a control word had to be used for more than one special word. The Dutch-English lexicon consisted of half of the words from the Dutch lexicon and half of the words from the English lexicon.

In the lexicons, each unique word occurred one time, two times, or four times, dependent on in which frequency group they were. A total list of the words used in the lexicons with their corresponding frequencies can be found in the Appendix.

In Table 1 I clearly present the number of words used for each of the basic simulations I will consider in chapter 3.

| | EngCog | EngFF | EngTE | EngCon | DutCog | DutFF | DutTE | DutCon | Total | TotalFreq |
|---|---|---|---|---|---|---|---|---|---|---|
| English monolingual | 40 | 40 | 80 | 128 | X | X | X | X | 288 | 665 |
| Dutch monolingual | X | X | X | X | 40 | 40 | 80 | 122 | 282 | 614 |
| Early Dutch-English bilingual | 20 | 20 | 40 | 69 | 20 | 20 | 40 | 69 | 298 | 655 |
| Late Dutch-English bilingual | 20 | 20 | 40 | 69 | 20 | 20 | 40 | 69 | 298 | 655 |
| Late English-Dutch bilingual | 20 | 20 | 40 | 69 | 20 | 20 | 40 | 69 | 298 | 655 |

**Table 1: The number of words used per simulation (EngCog = English cognates, EngFF = English false friends, EngTE = English translation equivalents, EngCon = English control words, DutCog = Dutch cognates, DutFF = Dutch false friends, DutTE = Dutch translation equivalents, DutCon = Dutch control words, Total = total number of unique words, TotalFreq = actual total number of words).**

Figure 1 shows the three layer feed forward network from Thomas that was used to learn the mappings between the orthographic and semantic codes. The network used in my research is the same, only the number of units in each layer differs. The reason for this difference is that I used the whole alphabet instead of only five letters and my lexicon exists of more words and thus needs more hidden units to map the codes. My network has an input layer of 105 units. Twenty-six units for each letter from the alphabet times four (word length) and at last one language coding unit. The output layer which is used to generate a meaning consists of 121 units. For the semantic feature unit 120 units are reserved and again the last one is the language coding unit. The cognates and translation equivalents were assigned the same meaning in both languages. The units in the input and output strings have a value of “0” or “1”. After some tests the number of units in the hidden layer is set on 400.

The learning algorithm used in my simulations is Leabra Hebbian Learning. Leabra stands for Local, Error-driven and Associative, Biologically Realistic Algorithm, and it implements a balance between Hebbian and error-driven learning on top of a biologically-based pointneuron activation function with inhibitory competition dynamics. More information about the model can be found in the report of  Erik Lormans.

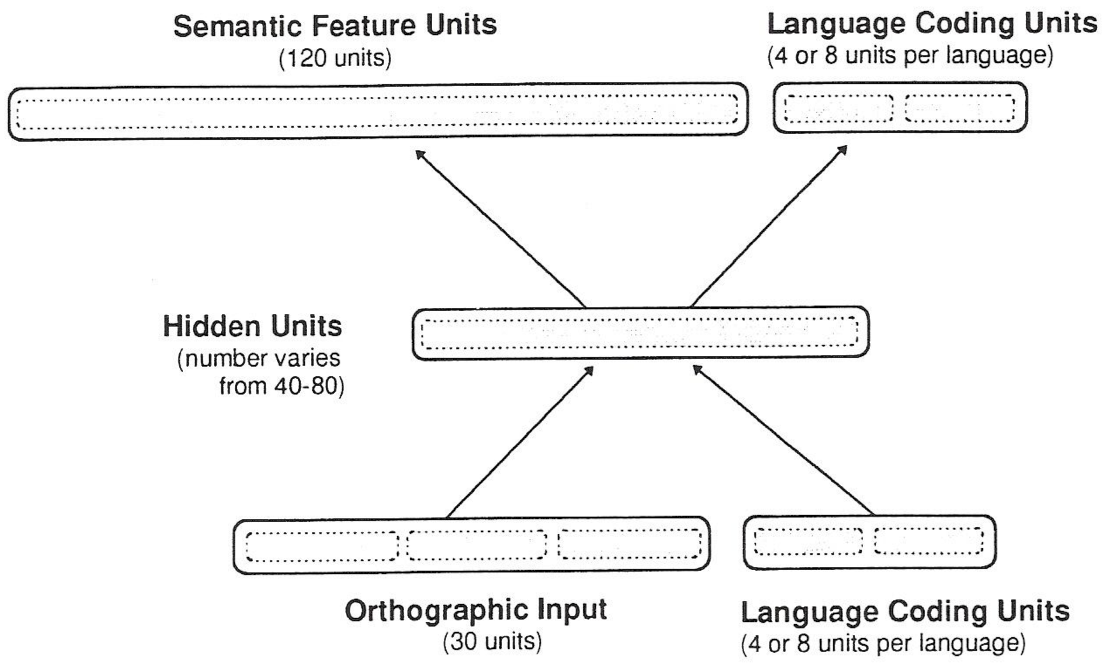


**Figure 1: The BSN-model of Thomas. The Extended BSN model used for my project had a similar structure, but different numbers of units: an input layer of 105 units, a hidden layer of 400 units and an output layer of 121 units.**

To measure the performance of the simulations and to compare them I will use the sum of the squares of the errors (SSE). This gives an indication of how wrong the outputs are during a simulation. I assume that a fast, easy learner makes less mistakes than a slow, difficult learner and thus has a lower SSE. Seidenberg and McClelland (1989) rationalize this assumption with the argument that, in a cascaded system, activation patterns that are asymptotically relatively clear (low in error) will reach a criterion level of clarity relatively quickly.

# 3. Basic simulations

### Simulation 1: Learning English as a first language

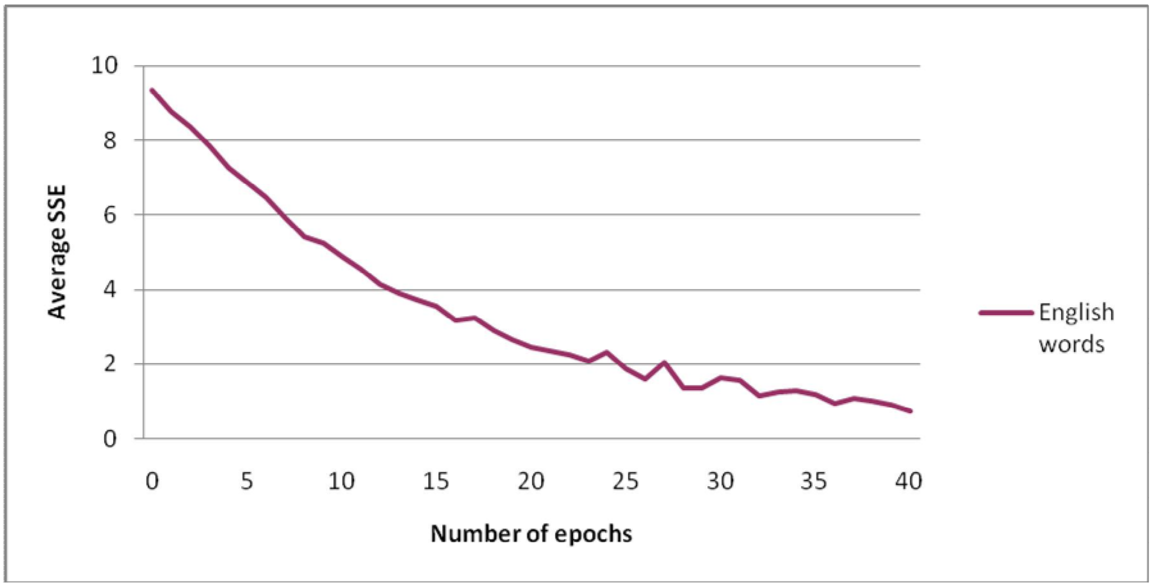


**Figure 2: Learning curve for an English monolingual.**

*Goal*

The goal of this simulation is to find out if the model is able to simulate the learning process of an English monolingual. This is a basic simulation that is indispensable further on in my research for comparisons with other monolingual and bilingual simulations.

*Simulation*

For this simulation I only considered the 128 control words in the English lexicon. The model needed 195 epochs to reach an average SSE of zero. The most interesting part of the learning curve is the beginning, so I only show the first forty epochs in the graph of Figure 2.

*Results and discussion*

Figure 2 shows that after forty epochs the average SSE reached a value below one. This indicates that my model is suitable for simulating the first-language acquisition of an English monolingual.

## Simulation 2: Learning Dutch as a first language

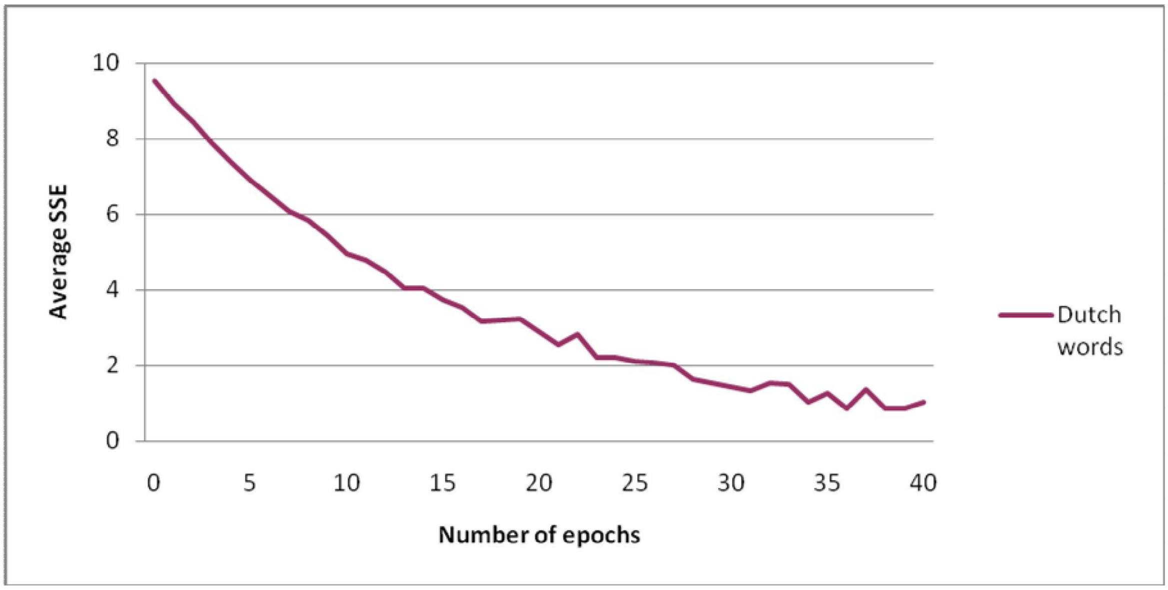


**Figure 3: Learning curve for a Dutch monolingual.**

*Goal*

The goal of this simulation is to find out if the model is able to simulate the learning process of a Dutch monolingual. Again, this is a basic simulation that will be used further on in my research for comparisons with other monolingual and bilingual simulations.

*Simulation*

For this simulation, I considered the 122 control words in the Dutch lexicon. The model needed 192 epochs to reach an average SSE of zero. The most interesting part of the learning curve is the beginning, so only the first forty epochs are shown in the graph of Figure 3.

*Results and discussion*

Figure 3 shows that after forty epochs the average SSE reached a value around one. This indicates that my model is suitable to simulate how Dutch monolinguals learn their mother tongue.

### Simulation 3: Learning English and Dutch simultaneously as an early bilingual

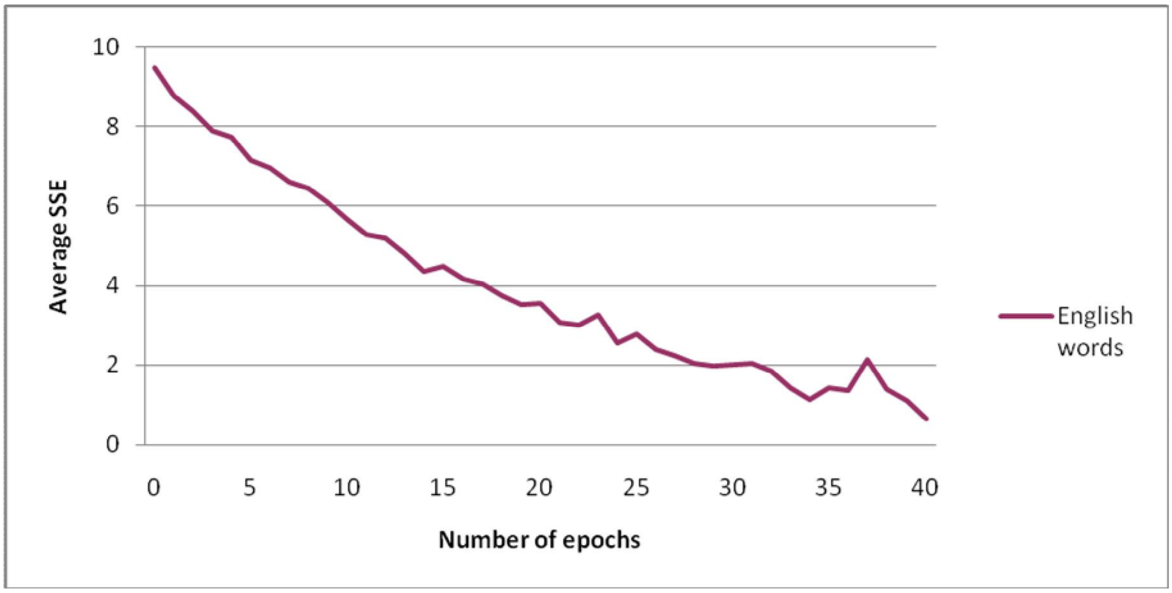


**Figure 4: Learning curve for how early Dutch-English bilinguals acquire English words.**

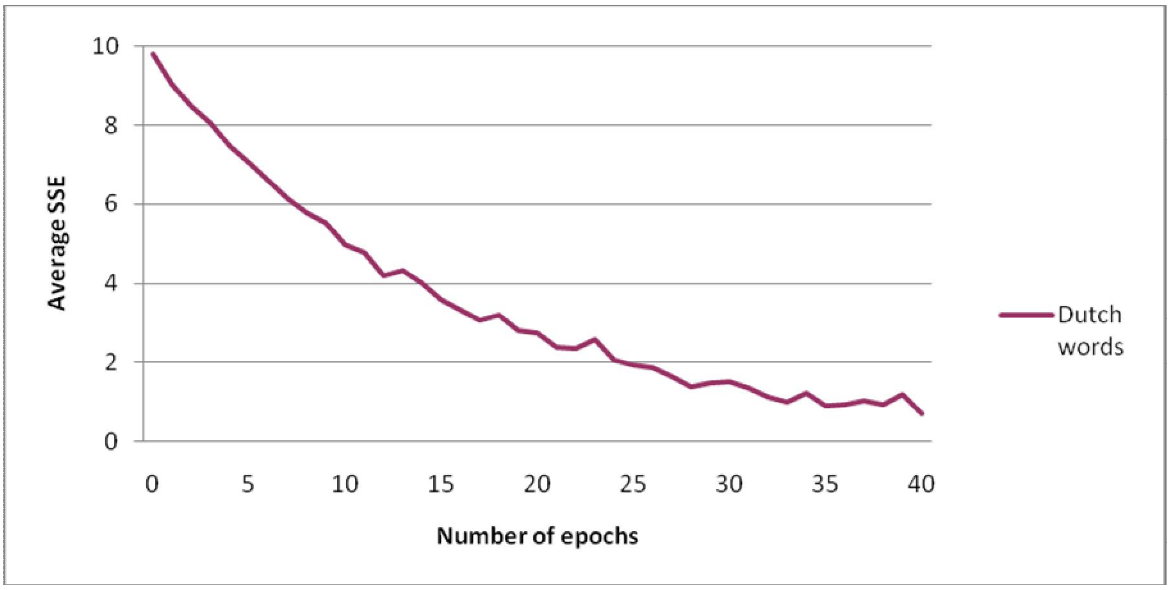


**Figure 5: Learning curve for how early Dutch-English bilinguals acquire Dutch words.**

*Goal*

The goal of this simulation is to find out if the model is able to simulate the learning process of an early Dutch-English bilingual. The simulation will be used for comparisons with other monolingual and bilingual simulations.

*Simulation*

For this simulation, I examined the 69 Dutch control words and the 69 English control words in the English-Dutch lexicon. The model needed 194 epochs to reach an average SSE of zero.

*Results and discussion*

Figure 4 and 5 both show that after forty epochs the average SSE reached a value around one. This indicates that my model is suitable for simulating the learning of an early Dutch-English bilingual.

## Simulation 4: Learning English as a late Dutch-English bilingual

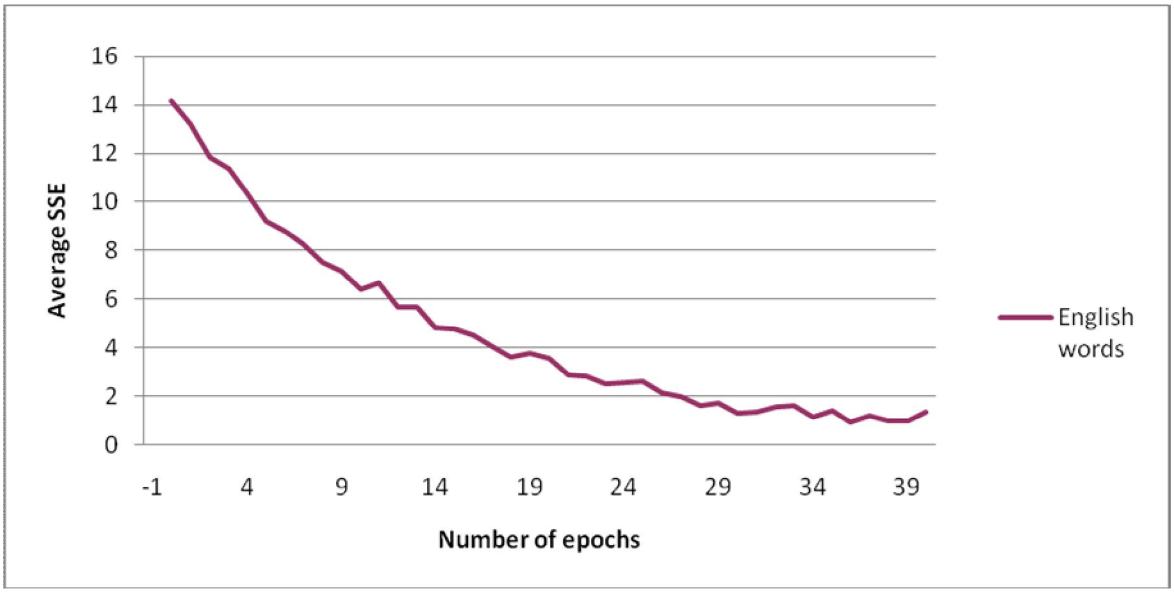


**Figure 6: Learning curve for how late Dutch-English bilinguals acquire English words.**

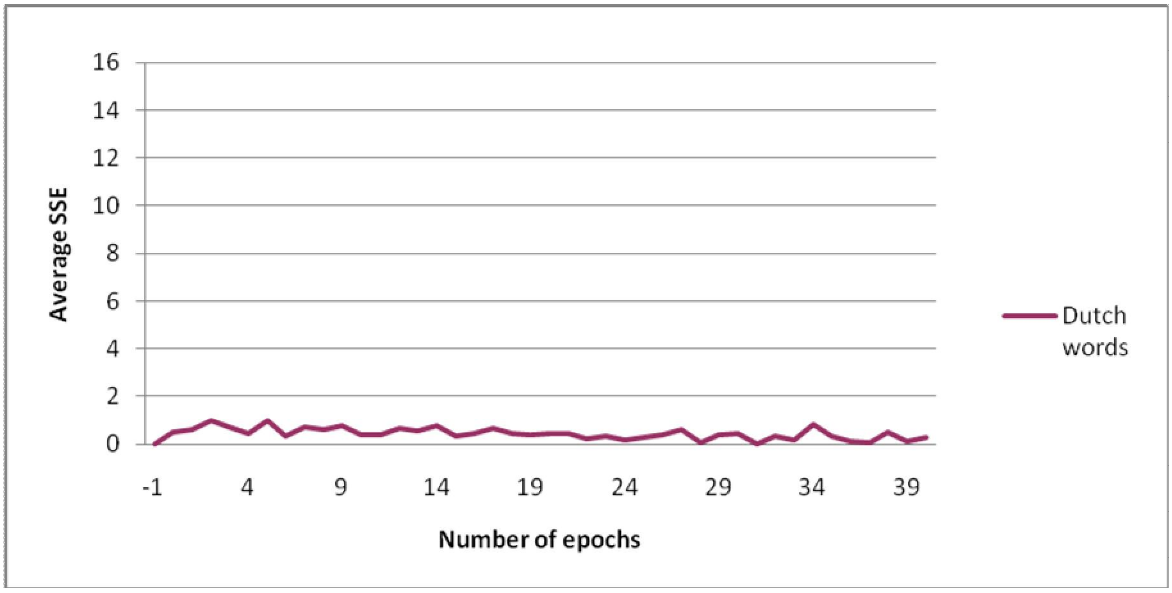


**Figure 7: Learning curve for how late Dutch-English bilinguals acquire Dutch words.**

*Goal*

This simulation mimics the English and Dutch word learning process of a late Dutch-English bilingual. This simulation will be used in subsequent research for comparisons with other monolingual and bilingual simulations.

*Simulation*

For this simulation I only considered the 69 Dutch control words and the 69 English control words in the English-Dutch lexicon. The model needed 319 epochs to reach an average SSE of zero. In Figure 6 and 7 I added epoch -1 to show that the Dutch words start with an average of 0, because these words were already learned by the model (as in Simulation 2). The only difference with simulation 2 is that in this simulation the model only needed to learn half of the Dutch words (69 words) to keep the number of Dutch words in the first and second part the same. The most interesting part of the learning curve is the onset stage of learning the second language, so I only show the first forty epochs in the graphs of Figure 6 and 7.

*Results and discussion*

Figure 6 shows that after forty epochs the average SSE of the English words reached a value around one. Figure 7 shows how the already learned Dutch words are affected by later learning of a second language. These two graphs indicate that my model can simulate word learning of a late Dutch-English bilingual.

**Simulation 5: Learning Dutch as a late English-Dutch bilingual**

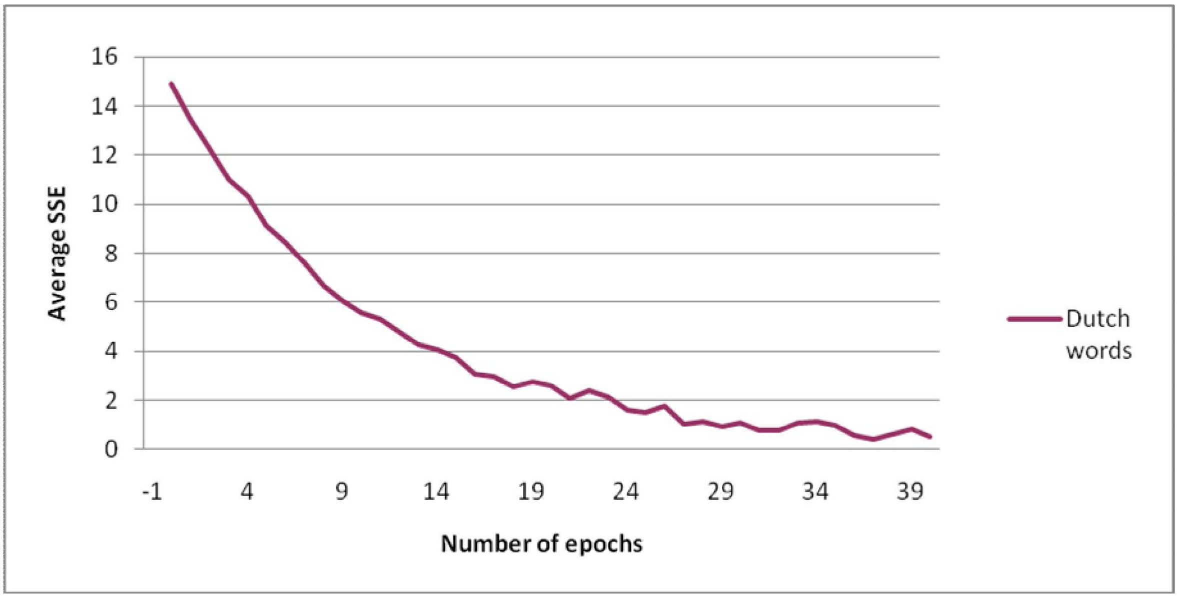


**Figure 8: Learning curve for how late English-Dutch bilinguals acquire Dutch words.**

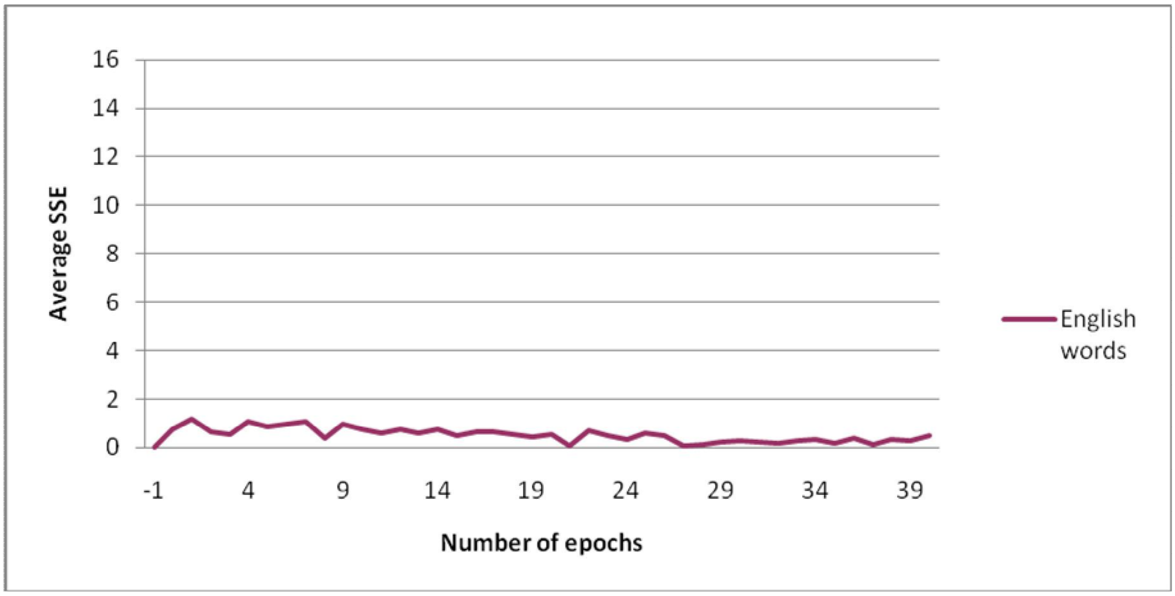


**Figure 9: Learning curve for how late English-Dutch bilinguals acquire English words.**

*Goal*

This simulation mimics the English and Dutch word learning process of a late English-Dutch bilingual. This simulation will be used in subsequent research for comparisons with other monolingual and bilingual simulations.

*Simulation*

For this simulation I only considered the 69 Dutch control words and the 69 English control words in the English-Dutch lexicon. The model needed 210 epochs to reach an average SSE of zero. In Figure 8 and 9 I added epoch -1 to show that the English words start with an average SSE of 0, because these words were already learned by the model (as in simulation 1). The only difference with simulation 1 is that in this simulation the model only needed to learn half of the English words (69 words) to keep the number of English words in the first and second part the same. The most interesting part of the learning curve is the onset stage of learning the second language, so I only show the first forty epochs in the graphs of Figure 8 and 9.

*Results and discussion*

Figure 8 shows that after forty epochs the average SSE of the Dutch words reached a value around one. Figure 9 shows how the already learned English words are affected by later learning of a second language. These two graphs indicate that my model can simulate word learning of a late English-Dutch bilingual.

# 4. Comparisons

**Comparison 1: English monolingual compared to a Dutch monolingual**

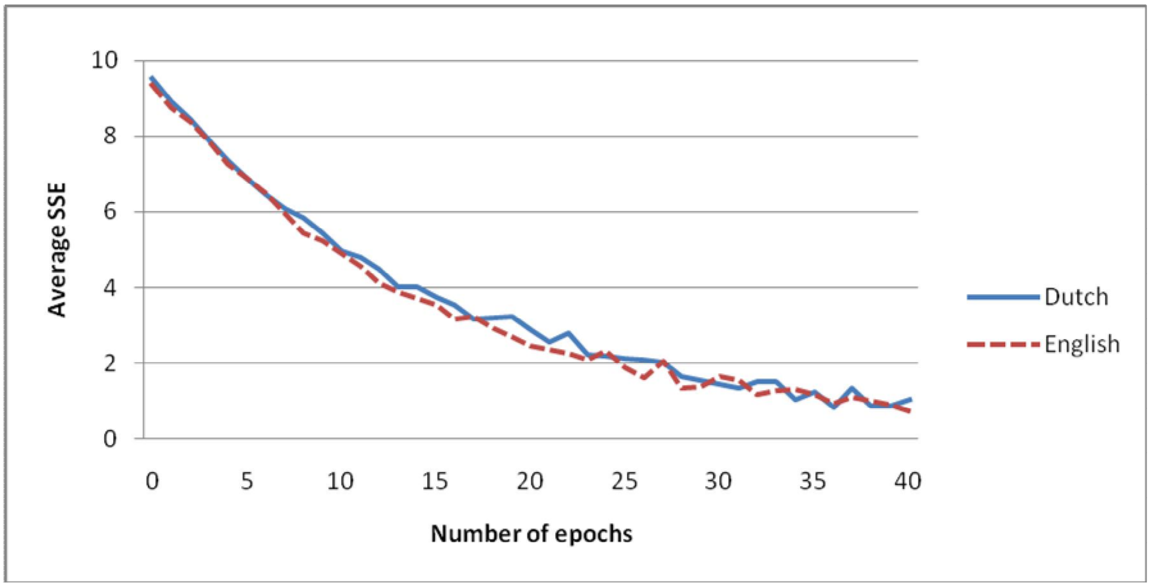


**Figure 10: Learning curves for how a Dutch monolingual acquires Dutch words and how an English monolingual acquires English words.**

*<u>Goal and hypotheses</u>*

I want to investigate if there is a difference between learning English as a monolingual or learning Dutch as a monolingual. Are words from one language more difficult to learn than from another language or is there no difference between them. I do not expect a difference between the two languages, because there also is no difference between monolinguals in England and in Holland regarding the age at which they can speak their language fluently.

*Simulation*

To answer this question I compared the two monolingual simulations mentioned in Chapter 3.

*Results and discussion*

Figure 10 shows that the Dutch and the English simulation follow the same learning curve. There is no difference in the quickness of the learning and there also is not a large difference in the error per epoch.

From this result we can conclude that our model is able to handle both languages Dutch and English the same way.

**Comparison 2: Learning English as a first language or as a second language after Dutch**

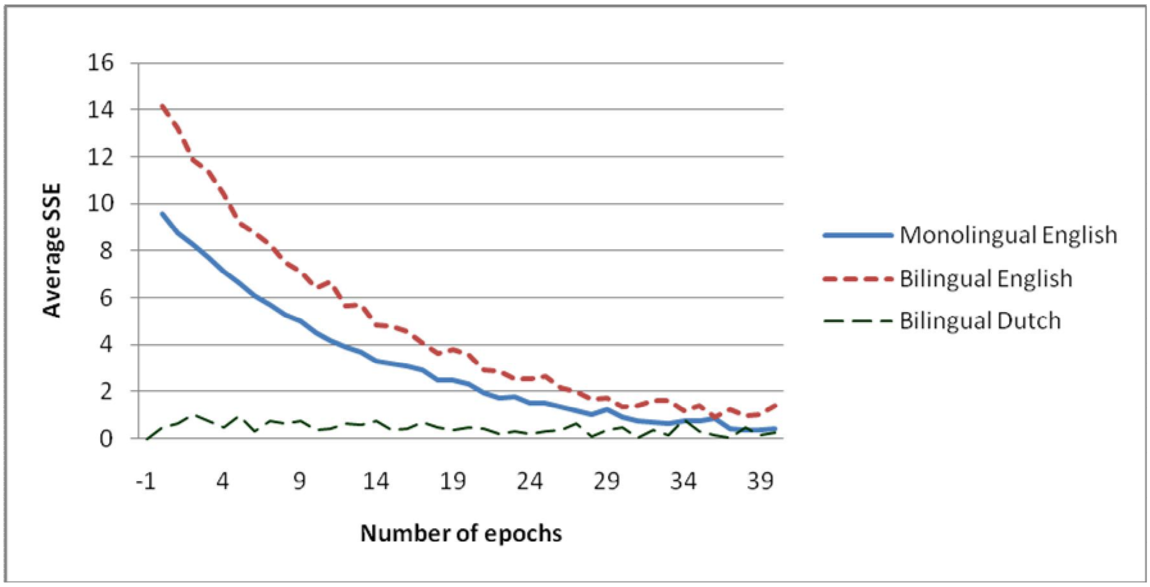


**Figure 11: Learning curves for how an English monolingual and a late Dutch-English bilingual acquire English words.**

*Goal and hypotheses*

The reason I did this simulation is to investigate if it helps to already know a language, before learning English, or that it is a disadvantage. It could be possible that already knowing a language helps learning another language, because of similarities between those languages. On the other hand could it also be possible that the first learned language makes it harder to learn a second language, because all the connections that have been made must be redirected. It is also interesting to see how the already learned Dutch words react on the addition of the English lexicon. Will their connections be disrupted or are they not affected by the learning of the English words?

*Simulation*

For this comparison I used the late Dutch-English bilingual simulation. I compared the learning curve from the words in this English monolingual simulation with the learning curve from the English words that were learned during the late Dutch-English bilingual simulation. I will also consider the learning curve of the Dutch words during the learning of the English words.

*Results and discussion*

Figure 11 shows that at the start of the learning process, the bilingual simulation begins with a higher SSE than the monolingual simulation. Apparently, the first learned language makes it harder for the bilingual simulation to learn the English words. For the rest of the epochs, the two learning curves approach each other more and more. This means that after a certain number of epochs the first language of the bilingual simulation is no longer a handicap. In the beginning the bilingual simulation makes more mistakes, but it learns better than the monolingual simulation. Figure 11 also indicates that the knowledge of the Dutch words is not affected by later learning of the second language.

**Comparison 3: Learning Dutch as a first language or as a second language after English**

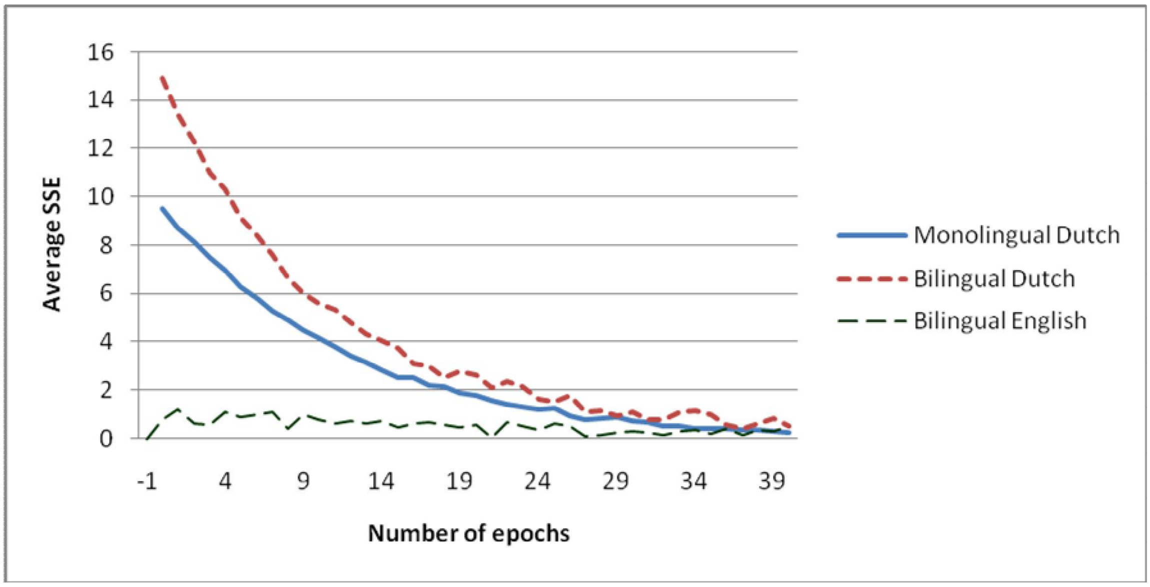


**Figure 12: Learning curves for how a Dutch monolingual and a late English-Dutch bilingual acquire Dutch words.**

*Goal and hypotheses*

With this simulation I investigate the size of the difference between learning Dutch as a monolingual or as a bilingual that already knows English. I also hope to see if it is an advantage to already know English when learning Dutch. It could also be a handicap to already know another language when learning a new language.

*Simulation*

For this comparison I used the late English-Dutch simulation. The Dutch words that were learned in this simulation were also used in the Dutch monolingual simulation. I compared the learning curve from the words in this Dutch monolingual simulation with the learning curve from the Dutch words that were learned during the late English-Dutch bilingual simulation. I will also have a look at the learning curve of the English words during the learning of the Dutch words.

*Results and discussion*

Figure 12 shows that the Dutch words that are learned during the bilingual simulation in the beginning have a higher SSE than the Dutch words learned in the monolingual simulation. As time progresses, the two learning curves come closer to eachother and the knowledge of the first language does not seem to be a handicap anymore.

In Figure 12 the learning curve of the already learned English words is shown from the moment the Dutch lexicon is added. The performance of the English words does not decrease when the Dutch words are added.

**Comparison 4: The difference between the late Dutch-English bilingual and the late English-Dutch bilingual**

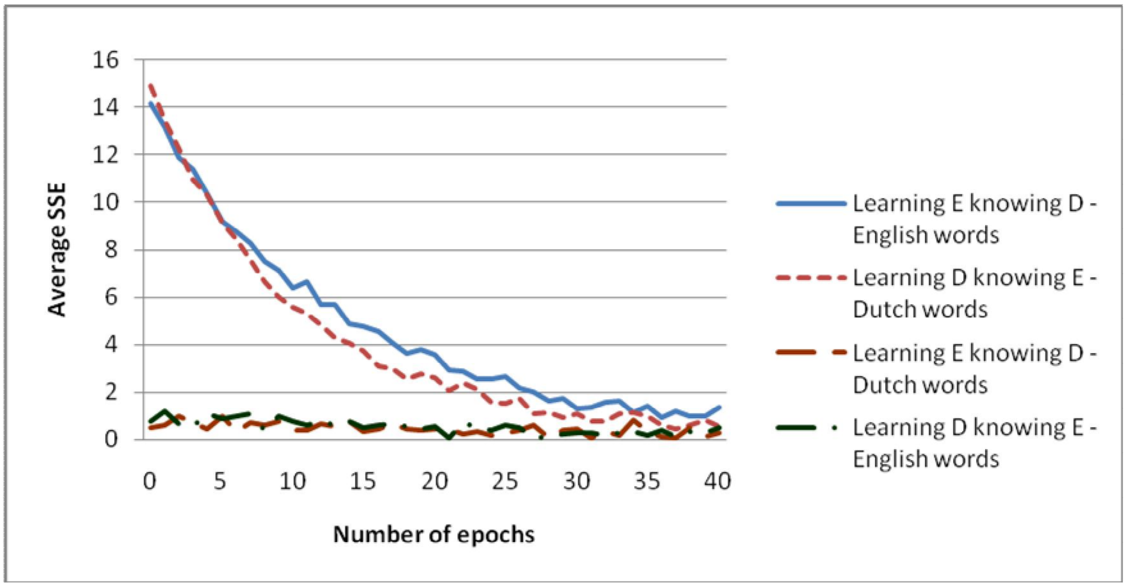


**Figure 13: late Learning curves for how late Dutch-English and late English-Dutch bilinguals acquire English and Dutch words.**

*Goal and hypotheses*

This comparison is interesting, because it shows if there is a difference in learning Dutch when knowing English or learning English when knowing Dutch. Earlier, it has already been shown that there is no difference between learning Dutch as a monolingual or learning English as a monolingual. In the recent comparison it could be possible that a difference in learning between both languages arises, due to the language that is already known.

*Simulation*

For this comparison I used the late Dutch-English bilingual simulation and the late English-Dutch bilingual simulation. I will have a look at the graphs of the second languages and the graphs of the native languages.

*Results and discussion*

At first Figure 13 shows that in both cases the already known first language is not affected by later learning of the second language. The learning curves of the second languages indicate that after five epochs a small difference originates. The native English speaker seems to learn the Dutch lexicon slightly easier than the native Dutch speaker learns the English lexicon,but this difference almost disappears after epoch thirty, so the learning curves are quite similar.

**Comparison 5: The difference between learning Dutch as a monolingual or as an early Dutch-English bilingual**

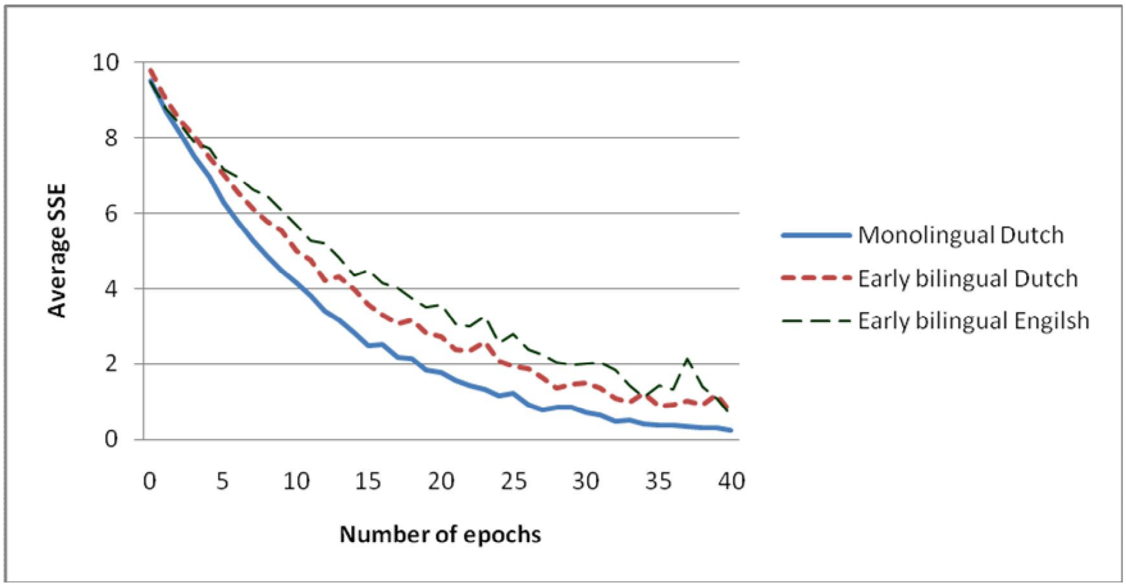


**Figure 14: Learning curves for how a Dutch monolingual and an early Dutch-English bilingual acquire Dutch words.**

*Goal and hypotheses*

With this comparison I want to find out if there is a difference between learning Dutch as a monolingual or as an early Dutch-English bilingual. I would expect that the early bilingual simulation is a bit slower with learning, because it has to learn two languages at the same time.

*Simulation*

For this comparison I used the early Dutch-English bilingual simulation. The Dutch words that were learned in this simulation were also used in the Dutch monolingual simulation (simulation 2). The only difference with simulation 2 is that in this simulation only half of the Dutch words are used to keep the number of Dutch words in the monolingual and the bilingual situation the same. I compared the learning curve of these Dutch words that were learned in the monolingual simulation to the learning curve of the Dutch words learned in the bilingual simulation. I also considered the English words in the early bilingual simulation.

*Results and discussion*

Figure 14 shows that the monolingual simulation learns the Dutch lexicon faster than the early bilingual simulation. This means that the early bilingual simulation suffers from the fact that it has to learn two languages at the same time. The learning curve of the English words is at some moments slower than the learning curve of the Dutch words, but there are also moments that they meet each other at the same level of average SSE, which means that the difference is small.

**Comparison 6: The difference between learning English as a monolingual or as an early Dutch-English bilingual**

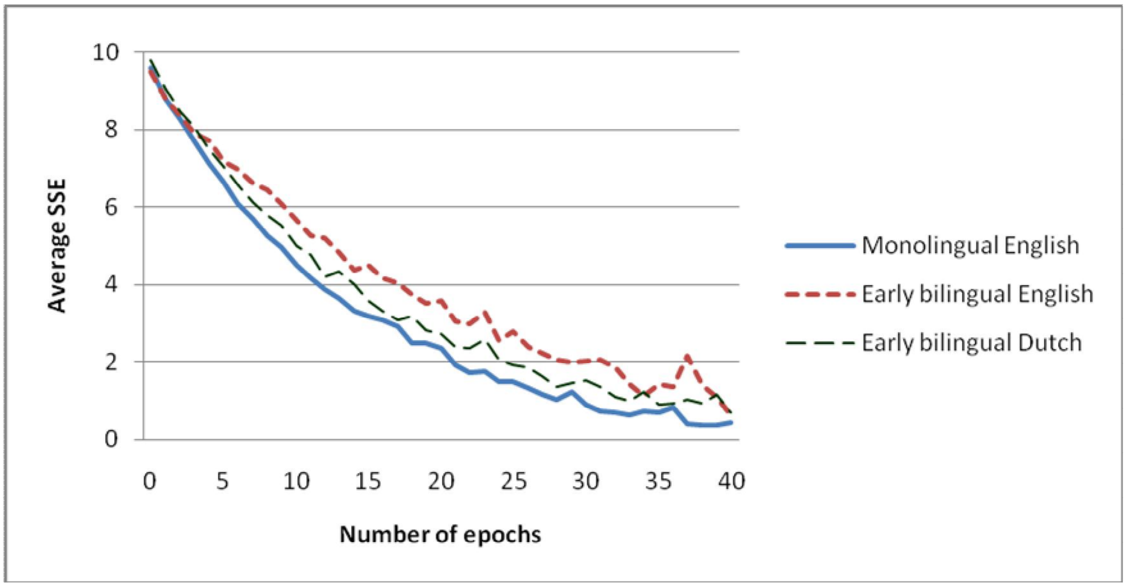


**Figure 15: Learning curves for how an English monolingual and an early Dutch-English bilingual acquire English words.**

*Goal and hypotheses*

This comparison is quite similar to the previous one, with the only difference being Dutch and English in opposite roles. It shows if there is a difference in learning English between a monolingual and an early bilingual. I expect to see the same type of effect as in the previous comparison. The early bilingual simulation will be slower than the monolingual simulation.

*Simulation*

For this comparison I again used the early Dutch-English bilingual simulation, but now the English words that were learned in this simulation were also used in the English monolingual simulation (simulation 1). The only difference with simulation 1 is that in this simulation only half of the English words are used to keep the number of English words in the monolingual and the bilingual situation the same. I compared  the learning curve of these English words that were learned in the monolingual simulation to the learning curve of the English words learned in the bilingual simulation. I also considered the Dutch words in the early bilingual simulation.

*Results and discussion*

Figure 15 shows that also in this comparison the English words of the monolingual simulation are learned faster than the English words in the early bilingual simulation. The learning curve of the Dutch words in the early bilingual simulation is a bit outstanding. It is positioned in the middle of the other two learning curves and the same type of effect with the Dutch words in the early bilingual was found in the previous comparison. Apparently, the Dutch words profit from the English words that are learned at the same time.

**Comparison 7: The difference between learning Dutch as an early Dutch-English bilingual or as a late English-Dutch bilingual**

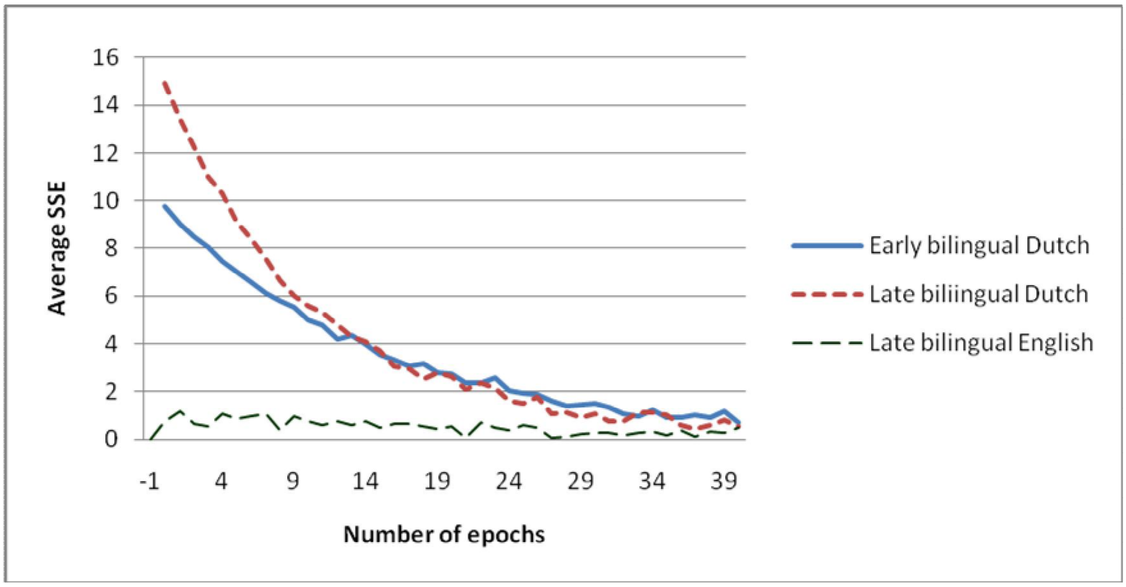


**Figure 16: Learning curves for how early Dutch-English and late English-Dutch bilinguals acquire Dutch words.**

*Goal and hypotheses*

With this comparison I want to find out if there is a difference between learning Dutch together with English or learning Dutch when the English lexicon is already known. It could be possible that it is an advantage to already know a language and that the second language profits from it. It could also be possible that the already known language makes it harder to learn a new language and that it is better to learn the two languages together.

*Simulation*

For this comparison I used the early Dutch-English bilingual simulation and the late English-Dutch simulation. I compared the learning curves of the Dutch words that were learned in both simulations. I also considered the English words in the late bilingual simulation.

*Results and discussion*

Figure 16 shows that there certainly is a difference between the two simulations. The late bilingual simulation seems to have a problem with learning the new Dutch words. Apparently, because of the already known English lexicon, the error in the beginning is higher than the error of the early bilingual simulation. However, the decrease in error is higher with the late bilingual simulation than with the early bilingual simulation. This leads to a decreasing difference between the lines and around epoch ten they follow the same curve. Both simulations end up with the same learning curve, but in the beginning the late bilingual simulation makes more mistakes.

**Comparison 8: The difference between learning English as an early Dutch-English bilingual or as a late Dutch-English bilingual**

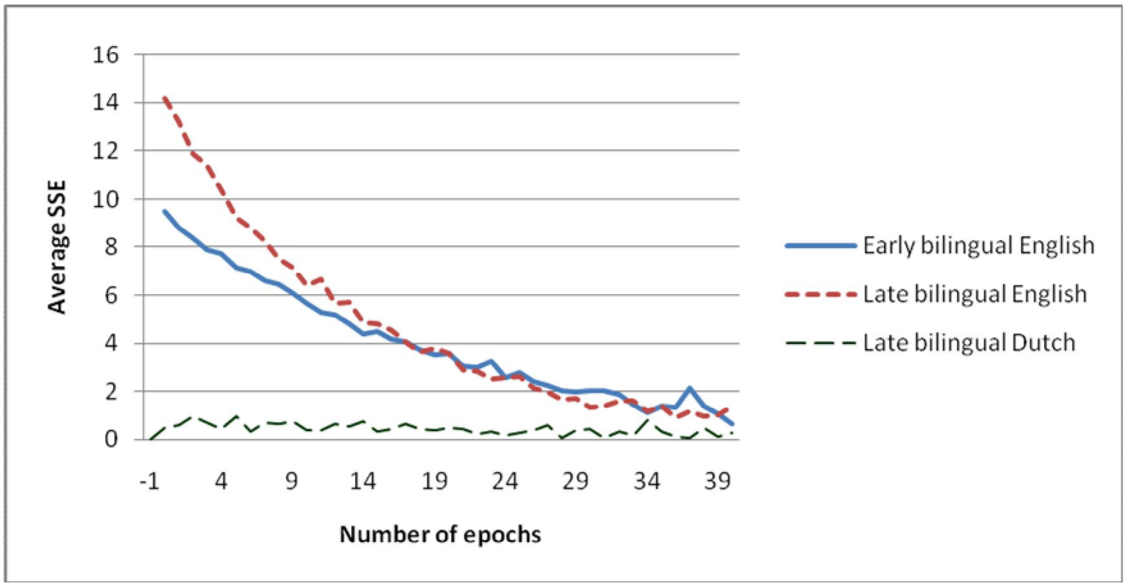


**Figure 17: Learning curves for how early Dutch-English and late Dutch-English bilinguals acquire English words.**

*Goal and hypotheses*

With this comparison I want to investigate if there is a difference between learning English as an early Dutch-English bilingual or as a late Dutch-English bilingual. Does it help to already know the Dutch lexicon or does it make learning English more difficult?

*Simulation*

For this comparison I used the early Dutch-English bilingual simulation and the late Dutch-English simulation. I compared the learning curves of the English words that were learned in both simulations. I also considered the Dutch words in the late bilingual simulation.

*Results and discussion*

Figure 17 shows that, just like the previous comparison I did, in the beginning, the error of the late bilingual simulation is higher than the error from the early bilingual simulation. From around epoch ten the lines come together and follow the same curve. In the beginning the early bilingual simulation has the advantage that both languages are new, but after a while this advantage is gone.

# 5. Special types of words

## Cognates, false friends, and translation equivalents

*Goal and hypotheses*

Now that the differences between monolinguals, early bilinguals and late bilinguals have been discussed, I want to consider some special types of words I have mentioned in the introduction: cognates, false friends, translation equivalents.

I think it is interesting to have look at how the learning curves of those special types of words differ from each other and how they differ from the control words.

I would expect that the cognates benefit from the fact that they occur in the English lexicon and in the Dutch lexicon while they are orthographically and semantically the same in both languages. If a cognate is already learned in one of the languages it is very easy for the model to also learn it in the other language, because the input (orthography) and output (semantics) are already known.

My expectation for the false friends is that they are more difficult to learn. When a false friend is already learned in one of the languages and is now being learned in the other language, the same input (orthography) has to be connected to another output (semantics). This will give some conflict and therefore will be harder to learn.

I think that the translation equivalents will profit from the fact that they are semantically the same in both languages. The model needs to connect different inputs to the same output which will make it easier to learn a translation equivalent that is already learned in one language.

*Simulation*

To find out if there is a difference in time needed to learn a word between the different types of words, I used three different simulations. In the first simulation both English and Dutch words are learned from the beginning (early Dutch-English bilingual). In the second and third simulations, first one language is learned and later the second language is learned (late Dutch-English bilingual and late English-Dutch bilingual).

*Results and discussion*

Figure 18 and 19 show that the different types of words have different learning curves in the early Dutch-English bilingual simulation. The false friends are harder to learn than the control words. The translation equivalents and the cognates are easier to learn than the control words. The cognates seem to profit from the fact that they semantically and orthographically occur in both languages. The translation equivalents only profit from the semantics that occur in both languages. The false friends are having trouble with the fact that the words that are orthographically the same in both languages have to be linked to different semantics.

Figure 20 shows that only the performance of the false friends is affected by later learning of a new language. The connections between the orthography and the semantics have to be rearranged. Figure 21 shows that the cognates profit the most. The new English cognates are immediately learned, because they are already learned with the Dutch lexicon. The translation equivalents also have a slight advantage, because their semantics are already learned in the Dutch lexicon. The false friends are slightly more difficult to learn than the control words. They suffer from the fact that their orthography is already connected with other semantics.

Figure 22 shows that also in this simulation the new cognates profit from the fact that they are already learned in the first lexicon. The translation equivalents also have an advantage, because their semantics have already been learned. Again, the false friends are more difficult to learn, because orthographically similar words have to be connected to different semantics. Figure 23 shows that the error of the false friends that were already learned is higher than the error from the other types of words. The new Dutch false friends make it harder to get the right semantics with the right orthography.

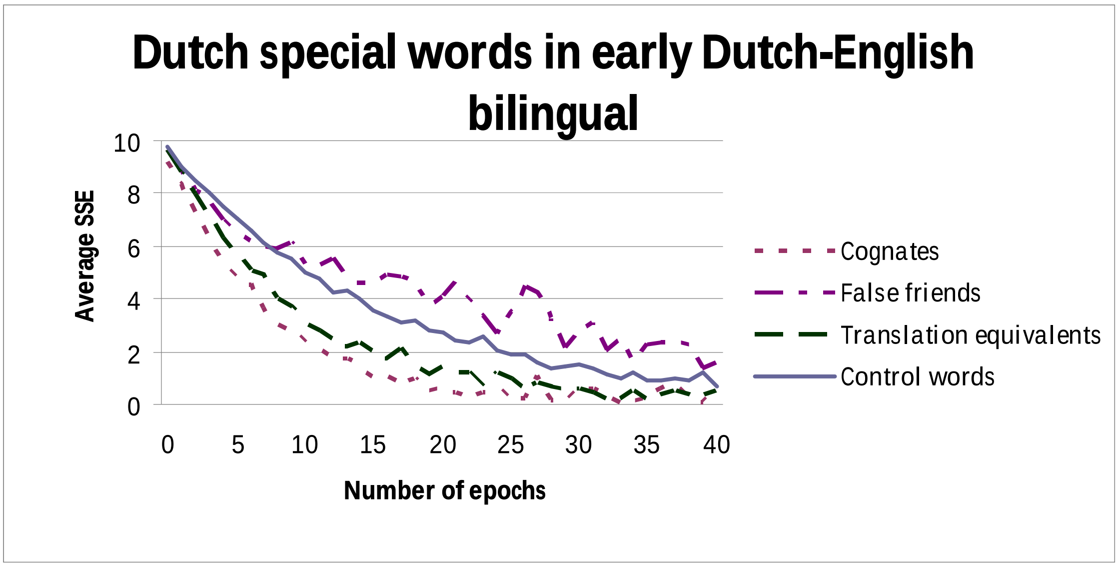


**Figure 18: Learning curves for how early Dutch-English bilinguals acquire Dutch cognates, false friends, translation equivalents, and control words.**

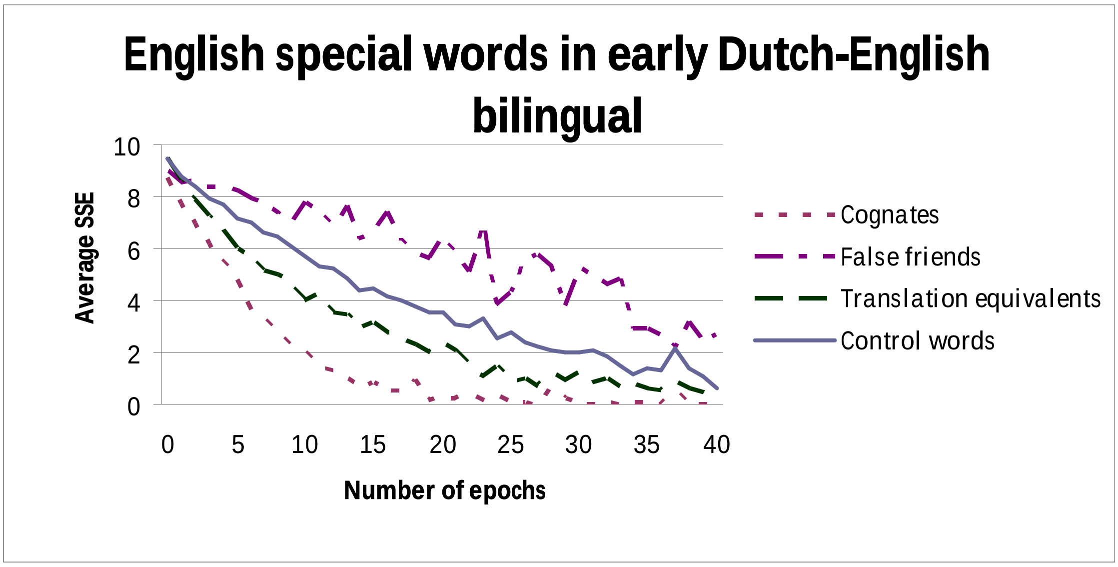


**Figure 19: Learning curves for how early Dutch-English bilinguals acquire English cognates, false friends, translation equivalents, and control words.**

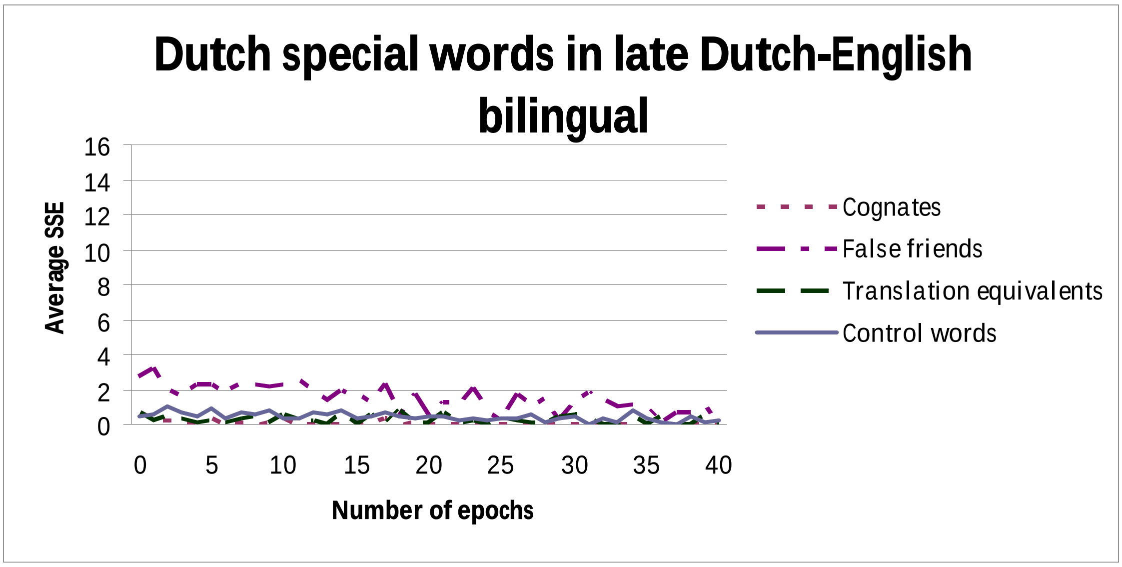


**Figure 20: Learning curves for how late Dutch-English bilinguals acquire Dutch cognates, false friends, translation equivalents, and control words.**

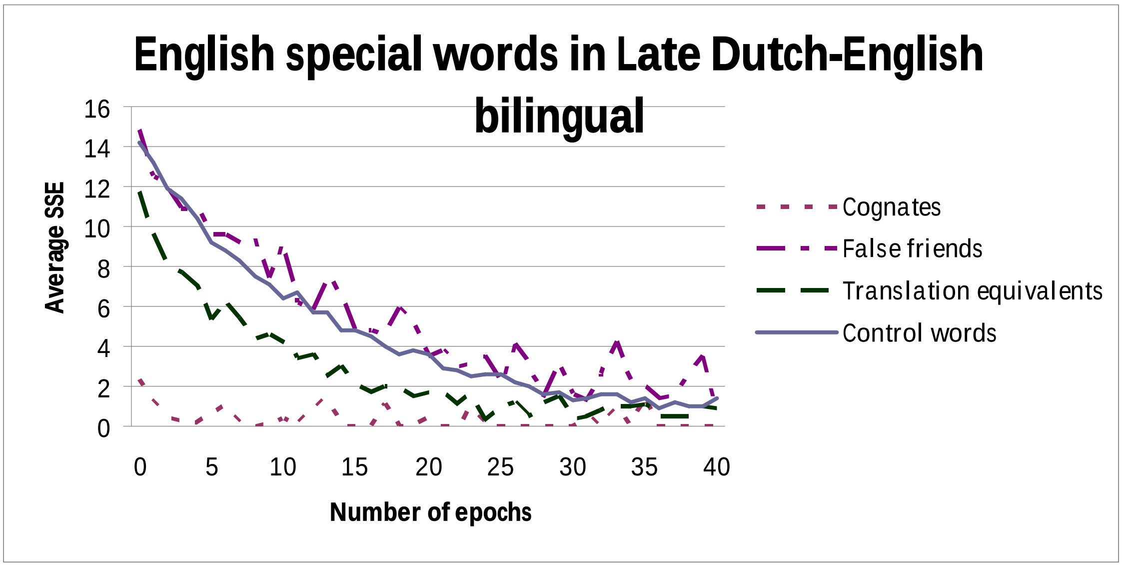


**Figure 21: Learning curves for how late Dutch-English bilinguals acquire English cognates, false friends, translation equivalents, and control words.**

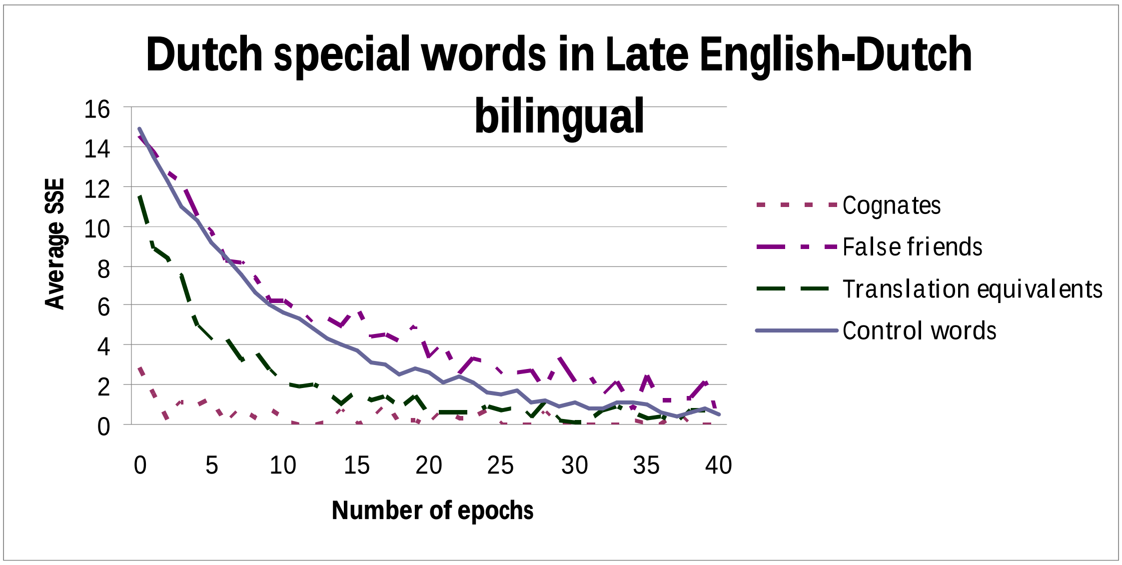


**Figure 22: Learning curves for how late English-Dutch bilinguals acquire Dutch cognates, false friends, translation equivalents, and control words.**

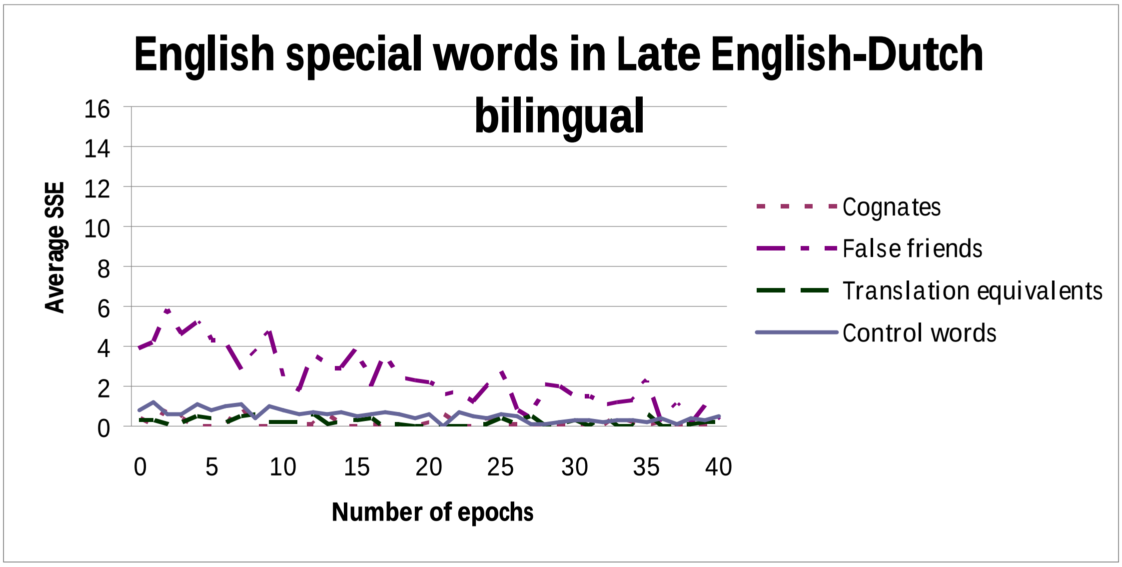


**Figure 23: Learning curves for how late English-Dutch bilinguals acquire English cognates, false friends, translation equivalents, and control words.**

## Translation equivalents

*Goal and hypotheses*

I divided the translation equivalents in four groups, where each group represented a level of overlap: translation equivalents with an overlap of zero (cage-kooi), translation equivalents with an overlap of one (deer-hert), translation equivalents with an overlap of two (dead-dood) and translation equivalents with an overlap of three (flag-vlag). This way cognates are equivalent to translation equivalents with an overlap of four.

When a person is learning two languages I would expect that the more overlap a translation equivalent has, the easier it is two learn this word. I also expect that the influence of having an overlap of zero or one is not very much, because a person does not notice if two words have only one letter in common.

*Simulation*

To find out if these hypotheses hold I again used the earlier mentioned three different simulations: early Dutch-English bilingual, late Dutch-English bilingual and late English-Dutch bilingual.

*Results and discussion*

Figure 24 and 25 show that the more overlap a set of translation equivalents has, the faster it is learned. Most of the overlap categories are learned faster than the control words which indicates that the similar semantics in both the lexicons fasten the learning process.
Figure 26 shows that the addition of a new lexicon from another language does not affect the knowledge over the already learned translation equivalents.

Figure 27 shows the effects that were found in Figure 24 and 25 in a stronger way. Translation equivalents with an overlap of three are learned fast and profit from the fact that they have an equivalent in the other language with the same semantics and almost the same orthography. The less overlap between the equivalents the harder it is for the model to learn them, but they are at least learned as fast as the control words.

Figure 28 and 29 show almost the same effects as Figure 26 and 27. The already learned translation equivalents do not get in trouble because of the new ones in the Dutch lexicon. The new Dutch translation equivalents are learned easier when the overlap is bigger.

They profit from the fact that their semantics have already been learned and that their orthography partly already has been learned.

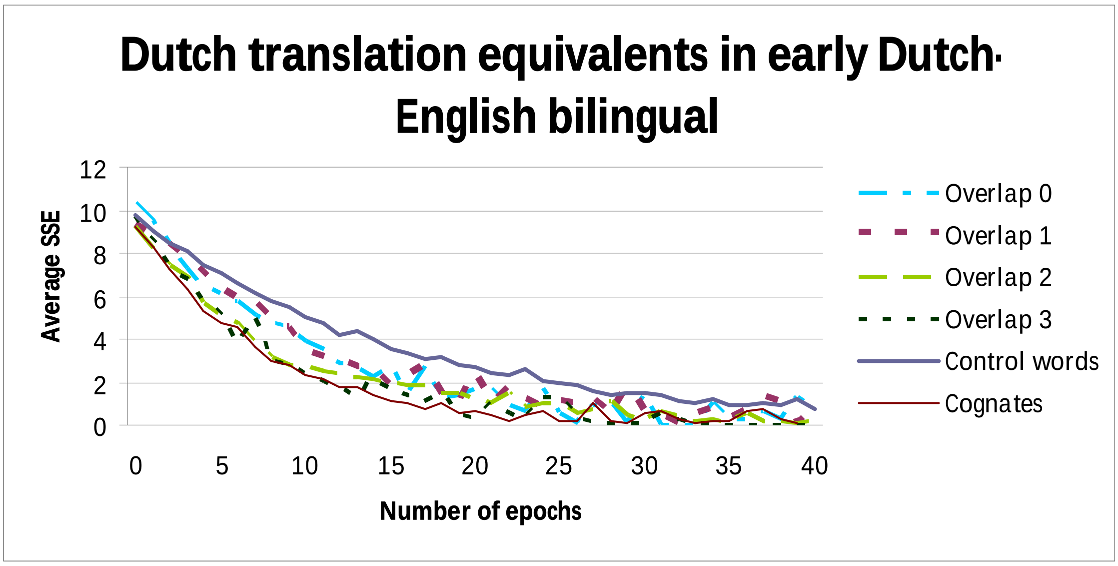


**Figure 24: Learning curves for how early Dutch-English bilinguals acquire the different categories of Dutch translation equivalents.**

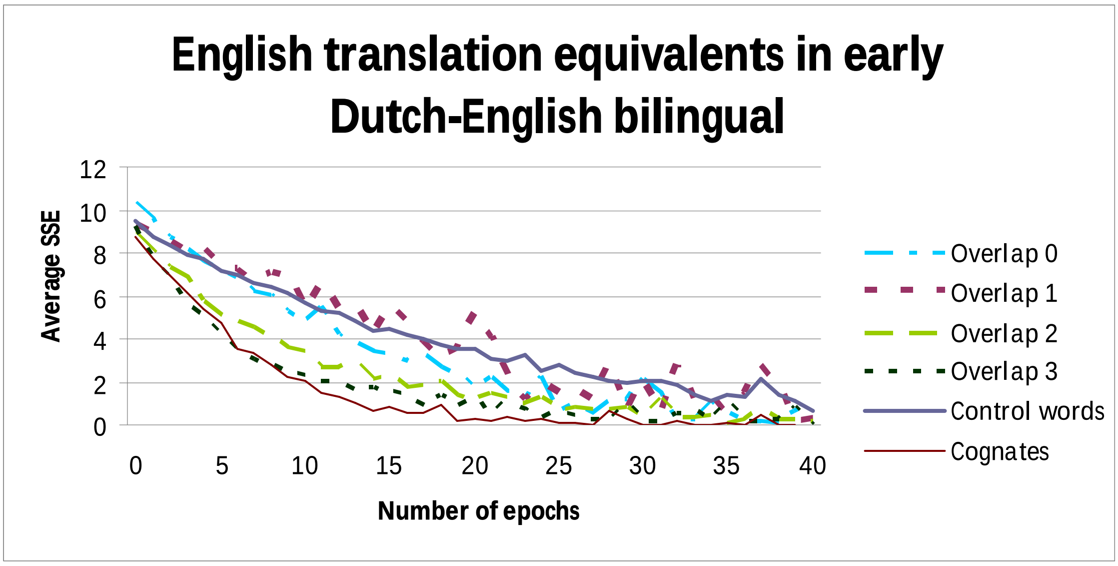


**Figure 25: Learning curves for how early Dutch-English bilinguals acquire the different categories of English translation equivalents.**

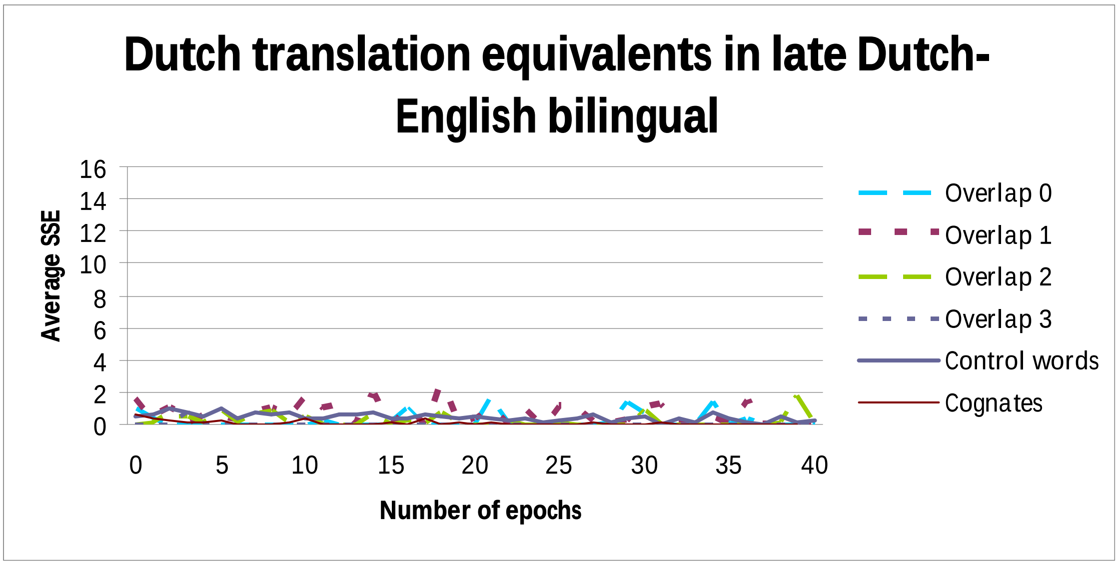


**Figure 26: Learning curves for how late Dutch-English bilinguals acquire the different categories of Dutch translation equivalents.**

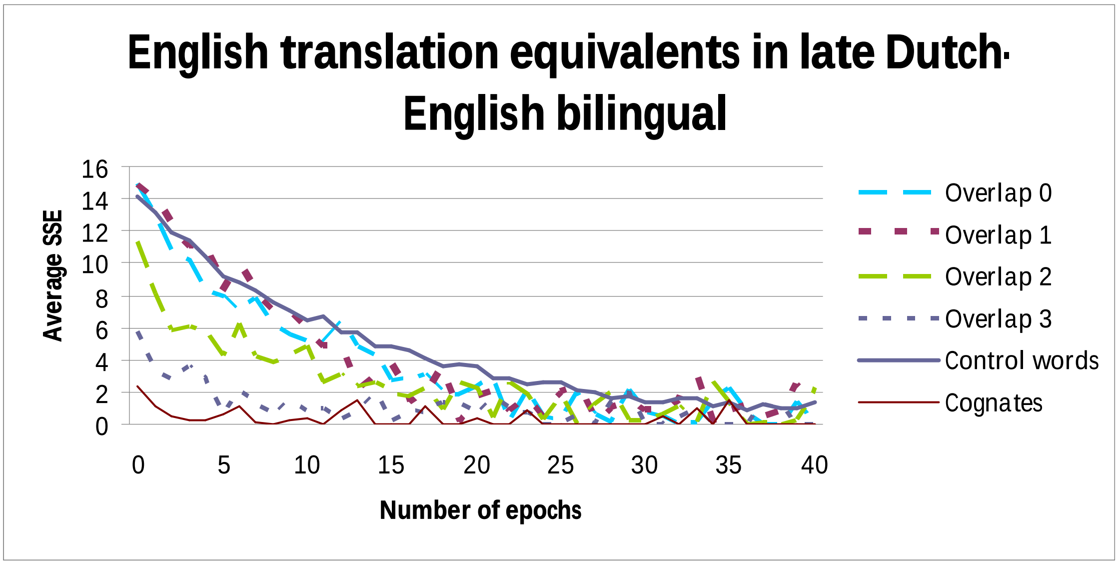


**Figure 27: Learning curves for how late Dutch-English bilinguals acquire the different categories of English translation equivalents.**

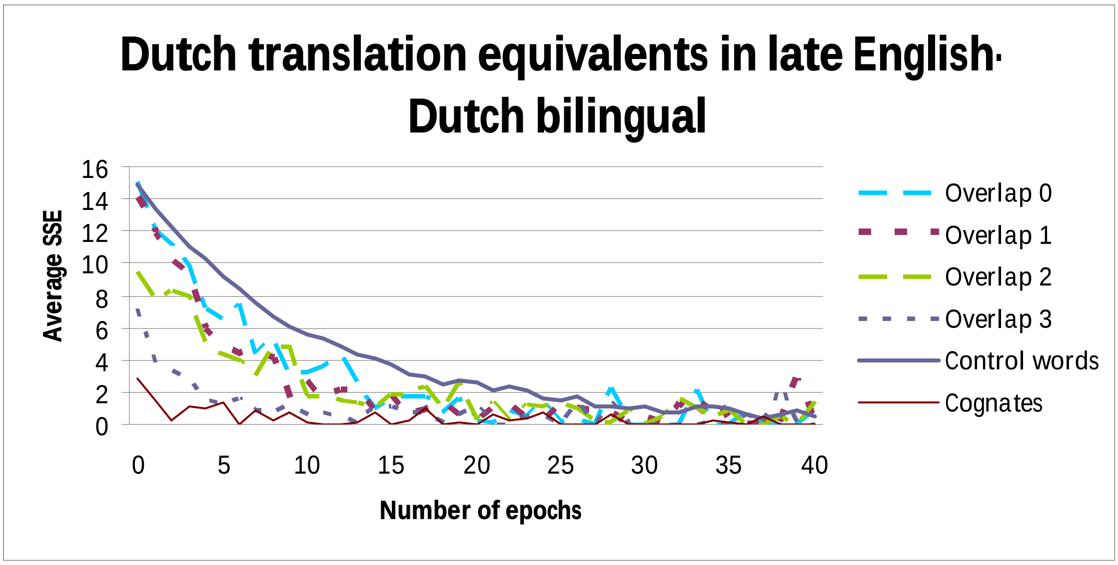


**Figure 28: Learning curves for how late English-Dutch bilinguals acquire the different categories of Dutch translation equivalents.**

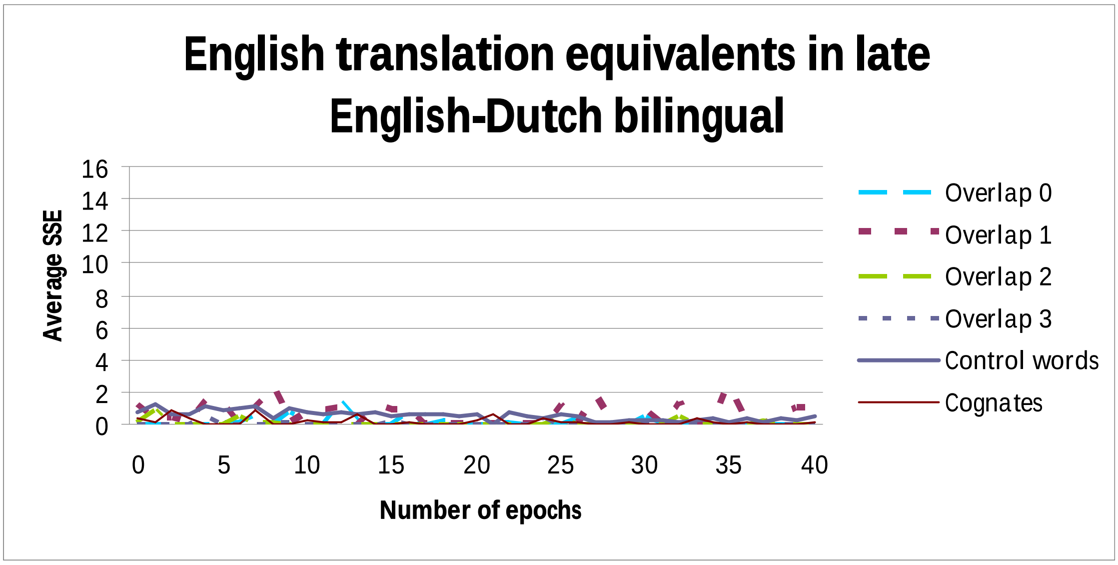


**Figure 29: Learning curves for how late English-Dutch bilinguals acquire the different categories of English translation equivalents.**

# 6. Word-frequency effect

*Goal and hypotheses*

The goal of this simulation was to get some more insight into the effects of the frequency of a word. I would expect that the more times a word is shown to the model, the faster it will be able to learn it. A word that is not shown very often will be inhibited by the words that are more frequent and therefore will be harder to learn.

*Simulation*

To see how the different frequency groups act, I used two simulations: the English monolingual simulation and the Dutch monolingual simulation.

*Results and discussion*

As shown in Figure 30 and 31, the frequency of a word does have a clear effect. The low-frequency words (Freq 1) are the most difficult to learn for the model and the high-frequency words (Freq 4) are the easiest to learn. The line of the middle-frequency words (Freq 2) is positioned in the middle of those two. The more a word is shown to the model the easier it will be learned.

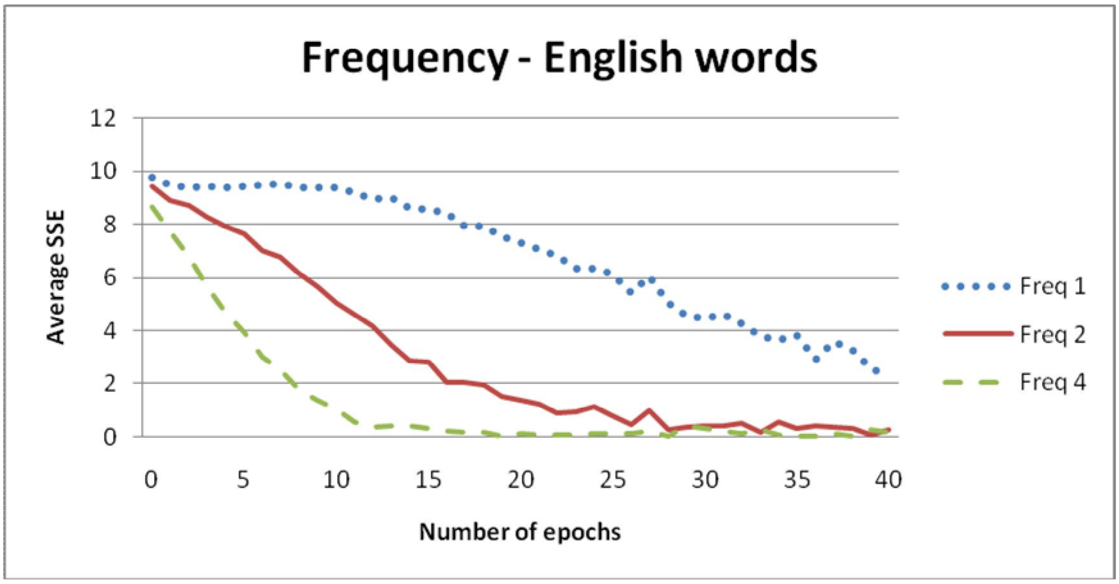


**Figure 30: Different frequency groups of English words compared.**

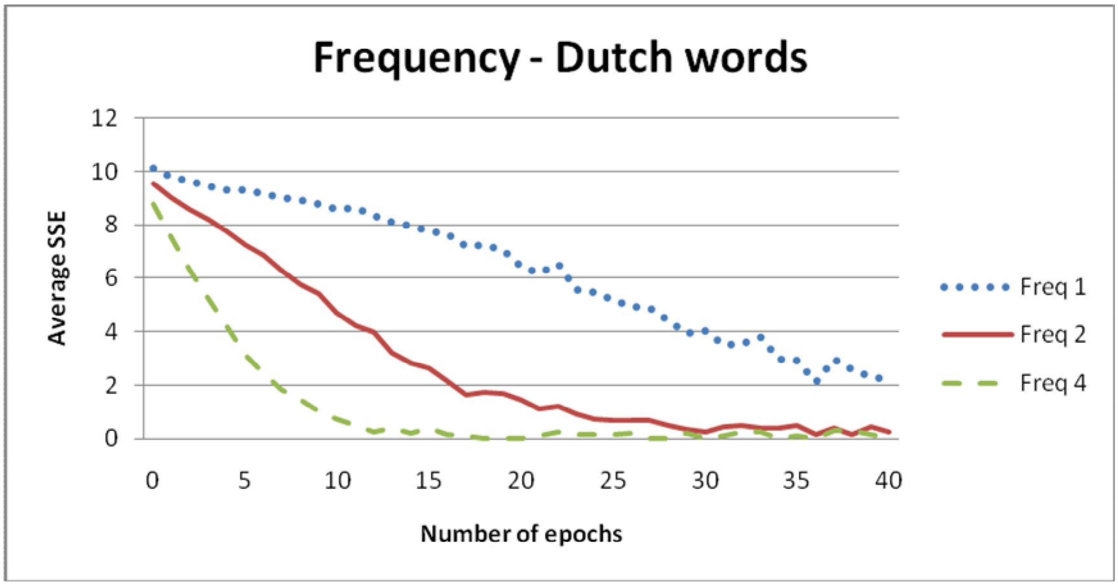


**Figure 31: Different frequency groups of Dutch words compared.**

# 7. General Discussion

In this research I simulated the acquisition of a second language by early and late bilinguals to explore any differences between these processes. Also, the learning processes of special types of words were analyzed and compared with empirical findings. I used a distributed connectionist model that was based on the network of Erik Lormans, who extended the BSN model of Thomas. The Dutch and English lexicons that were used consisted of cognates, false friends, translation equivalents, and control words. These words could have a high, middle, or low frequency. At first I did some comparisons between the monolingual, the early bilingual, and the late bilingual simulations. Next I compared the learning processes of the special types of words and at last I considered the learning processes of the three frequency groups.

The comparisons between the different types of bilinguals give us some general results. I found that the error rates for words of the native language of a late bilingual were not affected by the acquisition of a new language. In addition to this, the graphs show that the learning of the new language by a late bilingual is slowed down because of the fact that they possess a native language. Compared to the monolinguals this slowing down takes on longer than compared to the early bilingual simulations where this slowing down is only a short period. However, for all simulations hold, after forty epochs their handicap is conquered and their error level is the same.

These results suggest that it is very well possible to learn a second language after you have already acquired your native language. Just take in mind that in the beginning more errors will be made, but this decreases over time. The native language also will not be affected during this learning period. The results also contribute to the notion of the language-nonselective nature of the bilingual word recognition system. The native language of a late bilingual interferes with and slows down the learning of a new language. The same holds for an early bilingual where the two new languages interfere with each other and thus slow down each other. The learning process of a monolingual is not interfered by a second language and is therefore faster than the late and early bilinguals.

Still, we have to realize that early bilinguals do have an advantage. When a language is acquired via early bilingual processing, a second language has also been learned. So an early bilingual is slower when only one language is considered, but it is faster when two languages are considered.

My results are partly in line with the findings of Femke Haga, who did the same comparisons with a localist connectionist model. She also found that the learning of a second language as a late bilingual is slower than learning this language as a monolingual. The same holds for learning the second language as an early bilingual. In contradiction with my results is the finding that the native language in her simulations was affected by later learning of a second language.

The simulations with special types of words also provide some interesting results. The cognates are the easiest type of words to learn in early and late bilingual simulations. In the late bilingual simulations they even are almost immediately acquired. It is clear that they profit from the already available mappings between orthography and semantics. This is in line with the reaction time findings of Dijkstra et al. (2010) where cognates also had an advantage during processing (cognate facilitation effect) and provides evidence for language-nonselective access.

The translation equivalents are also easier learned than the control words, but still more difficult than the cognates. This makes sense, because the cognates actually are translation equivalents with an overlap of four letters (in four-letter words). The less overlap a translation equivalent has, the less easier it is to learn it. Just like the cognates they profit from the already available mappings between semantics and orthography. This was also found by Dijkstra et al., who talked about the similarity effect, where similar translation equivalents were recognized faster than translation equivalents with no similarity at all. This again contributes to the notion of the language-nonselective nature of the word recognition system.

False friends are the most difficult to learn. The results show that the learning of the false friends in the native language is affected by the addition of the false friends from the new second language. Mappings between orthography and semantics have to be rearranged and this results in an increased amount of errors. In the bilingual simulations we also see that the false friends overall are more difficult to learn than the control words. Dijkstra et al. also found this interference effect in their experiments which slowed down the reactions when a lexical decision task was done where words from one language were treated as non-words. Apparently, words from one language are not accessed independently from the other language.

Further research could be done on the comparison of the localist connectionist models with the distributed connectionist models. Let them acquire the same lexicons and try to equalize as many important elements as possible. It would be interesting to investigate if the same results could be produced and if they both could mimic the same empirical findings.

It would also be interesting to use larger and different lexicons to make the results more generalizable. Can the model simulate learning processes of all sorts of lexicons or is it specialized to a certain type? And is it still possible to produce learning curves that are in line with empirical findings under those circumstances?

To conclude, I would say that the extended BSN model is able to mimic the reaction time effects found in empirical studies. The cognates had an advantage during learning (cognate facilitation effect), similar translation equivalents were learned faster than translation equivalents with no similarity at all (similarity effect), the recognition of false friends was slowed down during the learning process (interference effect), and high frequency words are learned faster than low frequency words (word-frequency effect).

This research also showed that the process of learning a second language is slowed down by the knowledge of a native language. This contributes to the notion of the language-nonselective nature of the word recognition system.

# References

Baayen, H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database (CD -ROM)*. Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium.

Dijkstra, A. (2009). The multilingual lexicon. In Sandra, D., Östman, J.-O., & Verschueren, J. (Eds.), Cognition and language use: Handbook of Pragmatics Highlights. Also Handbook of Pragmatics On-Line.

Dijkstra, A. and Van Heuven, W. (1998). The BIA model and bilingual word recognition. In Grainger, J. and Jacobs, A., editors, Localist Connectionist Approaches to Human Cognition, pages 189–225. Mahwah, NJ, USA: Lawrence Erlbaum Associates.

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, doi:10.1016/j.jml.2009.12.003

Dijkstra, T., van Jaarsveld, H., and ten Brinke, S. (1998) Interlingual homograph recognition: Effects of task demands and intermixing. Bilingualism, Language and Cognition, 1:51-66.

French, R. and Jacquet, M. (2004). Understanding bilingual memory: models and data. TRENDS in Cognitive Sciences, 8(2), 87-93.

Haga, F. (2010). Development in the Bilingual Interactive Activation Model: Simulations of L1 and L2 Acquisition. Unpublished Bachelor Thesis Artificial Intelligence, Radboud University Nijmegen.

Lewy, N. and Grosjean, F. A computational model of bilingual lexical access. Manuscript in preparation.

Lormans, E. (2010). Bilingual lexical representation, Extending the Bilingual Single Network. Unpublished Bachelor Thesis Artificial Intelligence, Radboud University Nijmegen.

Seidenberg, M. and McClelland, J. (1989). A Distributed, Developmental Model of Word Recognition and Naming. Psychological Review, Vol. 96, No. 4, 523-568.

Thomas, M. (1997a). Connectionist networks and knowledge representation: The case of bilingual lexical processing. PhD thesis, Oxford University.

Thomas, M. (1997b). Distributed representations and the bilingual lexicon: One store or two? In Bullinaria, J., G. D. and Houghton, G., editors, Proceedings of the 4th Annual Neural Computation and Psychology Workshop. Springer.

Thomas, M. (1998). Bilingualism and the Single route / Dual Route debate. Proceedings of the 20th Annual Conference of the Cognitive Science Society, pages 1061–1066.

Van Heuven, W. and Dijkstra, A. The Semantic, Orthographic, and Phonological Interactive Activation Model. [Working title].

Van Heuven, W., Dijkstra, A., and Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. Journal of Memory and Language, 39:458–483.

# Appendix

Cognates are orthographically and semantically the same in both languages.

| Cognates | | | | | |
|---|---|---|---|---|---|
| **English** | **Dutch** | **Freq Eng** | **Freq Dut** | **Freq Eng Lexicon** | **Freq Dut Lexicon** |
| band | band | 49 | 75 | 2 | 4 |
| cake | cake | 38 | 3 | 2 | 1 |
| cape | cape | 17 | 5 | 2 | 1 |
| cent | cent | 304 | 26 | 4 | 2 |
| chef | chef | 4 | 37 | 1 | 2 |
| coma | coma | 2 | 2 | 1 | 1 |
| dame | dame | 3 | 89 | 1 | 4 |
| flat | flat | 134 | 31 | 4 | 2 |
| gong | gong | 2 | 2 | 1 | 1 |
| hand | hand | 830 | 1028 | 4 | 4 |
| homo | homo | 15 | 9 | 2 | 2 |
| jeep | jeep | 10 | 10 | 2 | 2 |
| mama | mama | 10 | 43 | 2 | 2 |
| mild | mild | 30 | 21 | 2 | 2 |
| mini | mini | 3 | 14 | 1 | 2 |
| oven | oven | 21 | 12 | 2 | 2 |
| papa | papa | 6 | 40 | 1 | 2 |
| park | park | 99 | 38 | 4 | 2 |
| pass | pass | 298 | 2 | 4 | 1 |
| pier | pier | 8 | 3 | 2 | 1 |
| pint | pint | 15 | 3 | 2 | 1 |
| plan | plan | 316 | 201 | 4 | 4 |
| pose | pose | 34 | 3 | 2 | 1 |
| race | race | 112 | 3 | 4 | 1 |
| rank | rank | 37 | 4 | 2 | 1 |
| rock | rock | 138 | 2 | 4 | 1 |
| show | show | 545 | 6 | 4 | 1 |
| take | take | 1920 | 4 | 4 | 1 |
| term | term | 99 | 88 | 4 | 4 |
| test | test | 149 | 15 | 4 | 2 |
| tram | tram | 5 | 20 | 1 | 2 |
| trio | trio | 4 | 3 | 1 | 1 |
| vest | vest | 8 | 10 | 2 | 2 |
| west | west | 50 | 16 | 2 | 2 |
| wild | wild | 98 | 62 | 4 | 4 |
| wind | wind | 143 | 111 | 4 | 4 |
| wolf | wolf | 13 | 17 | 2 | 2 |
| worm | worm | 20 | 10 | 2 | 2 |
| yoga | yoga | 2 | 4 | 1 | 1 |
| zone | zone | 17 | 11 | 2 | 2 |

False friends are orthographically the same in both languages, but semantically different.

| False friends | | | | | |
|---|---|---|---|---|---|
| **English** | **Dutch** | **Freq Eng** | **Freq Dut** | **Freq Eng Lexicon** | **Freq Dut Lexicon** |
| arts | arts | 38 | 93 | 2 | 4 |
| auto | auto | 4 | 208 | 1 | 4 |
| boot | boot | 40 | 67 | 2 | 4 |

| | | | | | |
|---|---|---|---|---|---|
| bout | bout | 9 | 2 | 2 | 1 |
| colt | colt | 2 | 2 | 1 | 1 |
| coup | coup | 9 | 4 | 2 | 1 |
| dank | dank | 2 | 79 | 1 | 4 |
| drop | drop | 177 | 2 | 4 | 1 |
| gist | gist | 2 | 3 | 1 | 1 |
| gust | gust | 4 | 2 | 1 | 1 |
| hart | hart | 4 | 190 | 1 | 4 |
| have | have | 13579 | 9 | 4 | 2 |
| last | last | 85 | 72 | 4 | 4 |
| leer | leer | 2 | 43 | 1 | 2 |
| list | list | 113 | 6 | 4 | 1 |
| loom | loom | 12 | 8 | 2 | 2 |
| mare | mare | 4 | 2 | 1 | 1 |
| mate | mate | 30 | 155 | 2 | 4 |
| meet | meet | 328 | 8 | 4 | 2 |
| mess | mess | 34 | 2 | 2 | 1 |
| most | most | 48 | 4 | 2 | 1 |
| peel | peel | 12 | 5 | 2 | 1 |
| perk | perk | 2 | 5 | 1 | 1 |
| pest | pest | 9 | 14 | 2 | 2 |
| pond | pond | 20 | 15 | 2 | 2 |
| prop | prop | 24 | 8 | 2 | 2 |
| punt | punt | 3 | 209 | 1 | 4 |
| ramp | ramp | 6 | 25 | 1 | 2 |
| rein | rein | 5 | 8 | 1 | 2 |
| roof | roof | 60 | 2 | 4 | 1 |
| room | room | 539 | 5 | 4 | 1 |
| slop | slop | 3 | 2 | 1 | 1 |
| slot | slot | 9 | 72 | 2 | 4 |
| spin | spin | 29 | 9 | 2 | 2 |
| trap | trap | 51 | 116 | 2 | 4 |
| vast | vast | 68 | 332 | 4 | 4 |
| veer | veer | 3 | 18 | 1 | 2 |
| vent | vent | 6 | 40 | 1 | 2 |
| wand | wand | 2 | 44 | 1 | 2 |
| zoom | zoom | 2 | 4 | 1 | 1 |

Translation equivalents are semantically the same in both languages, but orthographically different. The following table shows the translation equivalents with an orthographic overlap of 0.

| **Translation equivalents with an overlap of 0** | | | | | |
|---|---|---|---|---|---|
| **English** | **Dutch** | **Freq Eng** | **Freq Dut** | **Freq Eng Lexicon** | **Freq Dut Lexicon** |
| airy | fris | 3 | 37 | 1 | 2 |
| bent | krom | 3 | 9 | 1 | 2 |
| cage | kooi | 19 | 23 | 2 | 2 |
| cock | haan | 15 | 17 | 2 | 2 |
| cute | slim | 3 | 26 | 1 | 2 |
| game | spel | 212 | 113 | 4 | 4 |
| glue | lijm | 6 | 7 | 1 | 1 |
| gulp | slok | 8 | 26 | 2 | 2 |
| haze | mist | 7 | 11 | 1 | 2 |
| hull | romp | 3 | 12 | 1 | 2 |
| hush | stil | 10 | 165 | 2 | 4 |

| | | | | | |
|---|---|---|---|---|---|
| idle | loos | 15 | 5 | 2 | 1 |
| jack | boer | 54 | 100 | 2 | 4 |
| jeer | hoon | 4 | 2 | 1 | 0 |
| slum | krot | 11 | 3 | 2 | 1 |
| tree | boom | 203 | 137 | 4 | 4 |
| walk | loop | 380 | 103 | 4 | 4 |
| wall | muur | 226 | 147 | 4 | 4 |
| well | bron | 932 | 64 | 4 | 4 |
| wipe | veeg | 39 | 4 | 2 | 1 |

The following table shows the translation equivalents with an orthographic overlap of 1.

| **Translation equivalents with an overlap of 1** | | | | | |
|---|---|---|---|---|---|
| **English** | **Dutch** | **Freq Eng** | **Freq Dut** | **Freq Eng Lexicon** | **Freq Dut Lexicon** |
| bold | fors | 15 | 19 | 2 | 2 |
| clay | klei | 21 | 8 | 2 | 2 |
| crap | poep | 6 | 3 | 1 | 1 |
| curl | krul | 24 | 12 | 2 | 2 |
| deer | hert | 12 | 7 | 2 | 1 |
| four | vier | 2 | 4 | 1 | 1 |
| hose | kous | 4 | 12 | 1 | 2 |
| hump | bult | 8 | 4 | 2 | 1 |
| lace | kant | 18 | 293 | 2 | 4 |
| leaf | blad | 87 | 114 | 4 | 4 |
| loot | buit | 8 | 12 | 2 | 2 |
| mesh | maas | 4 | 2 | 1 | 1 |
| odor | geur | 11 | 70 | 2 | 4 |
| only | enig | 1791 | 205 | 4 | 4 |
| pall | maat | 2 | 40 | 1 | 2 |
| peat | veen | 4 | 6 | 1 | 1 |
| pope | paus | 6 | 27 | 1 | 2 |
| rake | hark | 8 | 2 | 2 | 1 |
| swig | teug | 2 | 11 | 1 | 2 |
| wood | hout | 102 | 48 | 4 | 2 |

The following table shows the translation equivalents with an orthographic overlap of 2.

| **Translation equivalents with an overlap of 2** | | | | | |
|---|---|---|---|---|---|
| **English** | **Dutch** | **Freq Eng** | **Freq Dut** | **Freq Eng Lexicon** | **Freq Dut Lexicon** |
| dead | dood | 192 | 386 | 4 | 4 |
| deaf | doof | 10 | 9 | 2 | 2 |
| drum | trom | 22 | 2 | 2 | 1 |
| epic | epos | 3 | 2 | 1 | 1 |
| flax | vlas | 2 | 2 | 1 | 1 |
| glib | glad | 2 | 37 | 1 | 2 |
| lane | laan | 44 | 11 | 2 | 2 |
| late | laat | 677 | 735 | 4 | 4 |
| name | naam | 403 | 420 | 4 | 4 |
| note | noot | 181 | 19 | 4 | 2 |
| reed | riet | 11 | 14 | 2 | 2 |
| sail | zeil | 27 | 15 | 2 | 2 |
| slab | plak | 9 | 11 | 2 | 2 |
| slug | slak | 3 | 5 | 1 | 1 |
| soot | roet | 2 | 3 | 1 | 1 |
| tart | trut | 4 | 4 | 1 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| time | tijd | 1975 | 1084 | 4 | 4 |
| vase | vaas | 8 | 12 | 2 | 2 |
| ware | waar | 2 | 570 | 1 | 4 |
| wide | wijd | 137 | 52 | 4 | 2 |

The following table shows the translation equivalents with an orthographic overlap of 3.

| **Translation equivalents with an overlap of 3** | | | | | |
|---|---|---|---|---|---|
| **English** | **Dutch** | **Freq Eng** | **Freq Dut** | **Freq Eng Lexicon** | **Freq Dut Lexicon** |
| bear | beer | 129 | 22 | 4 | 2 |
| beer | bier | 51 | 64 | 2 | 4 |
| flag | vlag | 29 | 29 | 2 | 2 |
| fork | vork | 18 | 12 | 2 | 2 |
| gold | goud | 92 | 36 | 4 | 2 |
| good | goed | 1420 | 1965 | 4 | 4 |
| haul | haal | 16 | 61 | 2 | 1 |
| hook | hoek | 57 | 111 | 4 | 4 |
| kirk | kerk | 8 | 205 | 2 | 4 |
| lung | long | 25 | 20 | 2 | 2 |
| mark | merk | 103 | 14 | 4 | 2 |
| meal | maal | 92 | 114 | 4 | 4 |
| milk | melk | 115 | 51 | 4 | 2 |
| pair | paar | 72 | 491 | 4 | 4 |
| pear | peer | 7 | 10 | 1 | 2 |
| stem | stam | 31 | 30 | 1 | 1 |
| step | stap | 167 | 124 | 2 | 2 |
| wasp | wesp | 6 | 4 | 1 | 1 |
| wick | wiek | 3 | 3 | 1 | 1 |
| work | werk | 1219 | 571 | 4 | 4 |

**ENGLISH LEXICON USED IN THE SIMULATIONS**

40 cognates
40 false friends
20 translation equivalents with an overlap of 0
20 translation equivalents with an overlap of 1
20 translation equivalents with an overlap of 2
20 translation equivalents with an overlap of 3

COGNATES + CONTROL WORDS

| | |
|---|---|
| band | cure |
| cake | fist |
| cape | flaw |
| cent | full |
| chef | fuze |
| coma | guru |
| dame | gush |
| flat | lack |
| gong | lira |
| hand | move |
| homo | pine |
| jeep | puff |
| mama | skip |
| mild | snap |
| mini | snug |
| oven | text |
| papa | tier |
| park | tiny |
| pass | view |
| pier | womb |
| pint | yawn |
| plan | able |
| pose | bait |
| race | beat |
| rank | bore |
| rock | deny |
| show | face |
| take | just |
| term | pack |
| test | pick |
| tram | plum |
| trio | pore |
| vest | roam |
| west | twin |
| wild | urge |
| wind | vary |
| wolf | veil |
| worm | vice |
| yoga | vide |
| zone | visa |

FALSE FRIENDS + CONTROL WORDS

| | |
|---|---|
| arts | bake |
| auto | bask |
| boot | chap |
| bout | cite |
| colt | dace |
| coup | dart |
| dank | eddy |
| drop | fall |
| gist | fang |
| gust | flak |
| hart | flea |
| have | will |
| last | wing |
| leer | yarn |
| list | army |
| loom | bail |
| mare | beak |
| mate | bulk |
| meet | fine |
| mess | flee |
| most | gaol |
| peel | gild |
| perk | glen |
| pest | gulf |
| pond | hack |
| prop | harm |
| punt | hike |
| ramp | honk |
| rein | hood |
| roof | hurt |
| room | keep |
| slop | kite |
| slot | lash |
| spin | lawn |
| trap | melt |
| vast | news |
| veer | nick |
| vent | nova |
| wand | oboe |
| zoom | ogre |

TRANSLATION EQUIVALENTS - Overlap 0 + CONTROL WORDS

| | |
|---|---|
| airy | prim |
| bent | fuze |
| cage | puff |
| cock | rash |
| cute | heed |
| game | fear |
| glue | rift |
| gulp | stab |
| haze | rump |
| hull | slay |
| hush | tidy |

| | |
|---|---|
| idle | womb |
| jack | bore |
| jeer | dual |
| slum | axis |
| tree | cost |
| walk | draw |
| wall | else |
| well | also |
| wipe | pray |

TRANSLATION EQUIVALENTS - Overlap 1 + CONTROL WORDS

| | |
|---|---|
| bold | gene |
| clay | boon |
| crap | pout |
| curl | dish |
| deer | puff |
| four | guru |
| hose | tier |
| hump | slim |
| lace | spit |
| leaf | save |
| loot | stab |
| mesh | dual |
| odor | tidy |
| only | much |
| pall | weir |
| peat | rung |
| pope | gush |
| rake | womb |
| swig | vide |
| wood | drag |

TRANSLATION EQUIVALENTS - Overlap 2 + CONTROL WORDS

| | |
|---|---|
| dead | wife |
| deaf | vice |
| drum | acre |
| epic | rove |
| flax | lira |
| glib | pomp |
| lane | keen |
| late | seem |
| name | fish |
| note | pull |
| reed | knit |
| sail | grin |
| slab | pine |
| slug | tier |
| soot | pomp |
| tart | dual |
| time | back |
| vase | puff |
| ware | crux |
| wide | ball |

TRANSLATION EQUIVALENTS - Overlap 3 + CONTROL WORDS

| | |
|---|---|
| bear | soft |
| beer | grin |
| flag | luck |
| fork | vice |
| gold | drag |
| good | back |
| haul | visa |
| hook | earn |
| kirk | gene |
| lung | snap |
| mark | pour |
| meal | push |
| milk | push |
| pair | roll |
| pear | rift |
| stem | boon |
| step | wear |
| wasp | slay |
| wick | snug |
| work | give |

**DUTCH LEXICON USED IN THE SIMULATIONS**

40 cognates
40 false friends
20 translation equivalents with an overlap of 0
20 translation equivalents with an overlap of 1
20 translation equivalents with an overlap of 2
20 translation equivalents with an overlap of 3

COGNATES + CONTROL WORDS

| | |
|---|---|
| band | erin |
| cake | gier |
| cape | gift |
| cent | grof |
| chef | hemd |
| coma | kiwi |
| dame | kost |
| flat | mouw |
| gong | prat |
| hand | soms |
| homo | wieg |
| jeep | zuid |
| mama | heus |
| mild | juni |
| mini | kaak |
| oven | naar |
| papa | riks |
| park | seks |
| pass | shag |
| pier | smid |
| pint | snee |
| plan | vorm |
| pose | zalm |
| race | cola |
| rank | dito |
| rock | kaft |
| show | klik |
| take | liga |
| term | raad |
| test | raak |
| tram | roer |
| trio | rouw |
| vest | stal |
| west | doop |
| wild | fase |
| wind | fles |
| wolf | gade |
| worm | hall |
| yoga | hars |
| zone | heil |

FALSE FRIENDS + CONTROL WORDS

| | |
|---|---|
| arts | baan |
| auto | bang |
| boot | fijn |
| bout | fuif |
| colt | fuik |
| coup | fust |
| dank | geel |
| drop | guur |
| gist | haai |
| gust | halm |
| hart | iets |
| have | koek |
| last | reis |
| leer | sint |
| list | slet |
| loom | star |
| mare | sten |
| mate | dier |
| meet | ende |
| mess | etui |
| most | faam |
| peel | fiks |
| perk | fooi |
| pest | ford |
| pond | gaaf |
| prop | hees |
| punt | jong |
| ramp | juli |
| rein | kode |
| roof | kwik |
| room | luis |
| slop | nijd |
| slot | rode |
| spin | roes |
| trap | veld |
| vast | eeuw |
| veer | eraf |
| vent | hoed |
| wand | hoer |
| zoom | huls |

TRANSLATION EQUIVALENTS - Overlap 0 + CONTROL WORDS

| | |
|---|---|
| fris | hemd |
| krom | lauw |
| kooi | riks |
| haan | wieg |
| slim | nood |
| spel | dorp |
| lijm | hars |
| slok | mouw |
| mist | lomp |
| romp | nazi |
| stil | vorm |

| | |
|---|---|
| loos | gift |
| boer | erin |
| hoon | leep |
| krot | lans |
| boom | mede |
| loop | plek |
| muur | faze |
| bron | raad |
| veeg | spar |

TRANSLATION EQUIVALENTS - Overlap 1 + CONTROL WORDS

| | |
|---|---|
| fors | puin |
| klei | zuid |
| poep | wenk |
| krul | naar |
| hert | zalm |
| vier | hels |
| kous | stal |
| bult | sans |
| kant | orde |
| blad | fase |
| buit | gade |
| maas | neen |
| geur | erin |
| enig | oude |
| maat | gids |
| veen | erwt |
| paus | mouw |
| hark | prat |
| teug | drie |
| hout | klas |

TRANSLATION EQUIVALENTS - Overlap 2 + CONTROL WORDS

| | |
|---|---|
| dood | taal |
| doof | kruk |
| trom | keet |
| epos | klim |
| vlas | zulk |
| glad | kast |
| laan | doop |
| laat | soms |
| naam | kind |
| noot | hall |
| riet | heil |
| zeil | drie |
| plak | Inkt |
| slak | aura |
| roet | smid |
| trut | snee |
| tijd | keer |
| vaas | inkt |
| waar | heen |
| wijd | juni |

TRANSLATION EQUIVALENTS - Overlap 3 + CONTROL WORDS

| | |
|---|---|
| beer | kalm |
| bier | leuk |
| vlag | grof |
| vork | kaar |
| goud | kast |
| goed | maar |
| haal | worp |
| hoek | fase |
| kerk | ooit |
| long | zalf |
| merk | dief |
| maal | dorp |
| melk | boel |
| paar | heen |
| peer | kruk |
| stam | juni |
| stap | kost |
| wesp | aura |
| wiek | faam |
| werk | soms |

**ENGLISH-DUTCH LEXICON USED IN THE SIMULATIONS**

20 Dutch cognates + 20 English cognates
20 Dutch false friends + 20 English false friends
10 Dutch translation equivalents with an overlap of 0 + 10 English translation equivalents with an overlap of 0
10 Dutch translation equivalents with an overlap of 1 + 10 English translation equivalents with an overlap of 1
10 Dutch translation equivalents with an overlap of 2 + 10 English translation equivalents with an overlap of 2
10 Dutch translation equivalents with an overlap of 3 +10 English translation equivalents with an overlap of 3

DUTCH COGNATES + CONTROL WORDS

| | |
|---|---|
| band | erin |
| cape | gift |
| chef | hemd |
| dame | kost |
| gong | prat |
| homo | wieg |
| mama | heus |
| mini | kaak |
| papa | riks |
| pass | shag |
| pint | snee |
| pose | zalm |
| rank | dito |
| show | klik |
| term | raad |
| tram | roer |
| vest | stal |
| wild | fase |
| wolf | gade |
| yoga | hars |

DUTCH FALSE FRIENDS + CONTROL WORDS

| | |
|---|---|
| arts | baan |
| boot | fijn |
| colt | fuik |
| dank | geel |
| gist | haai |
| hart | iets |
| last | reis |
| list | slet |
| mare | sten |
| meet | ende |
| most | faam |
| perk | fooi |
| pond | gaaf |
| punt | jong |
| rein | kode |
| room | luis |
| slot | rode |
| trap | veld |
| veer | eraf |
| wand | hoer |

DUTCH TRANSLATION EQUIVALENTS - Overlap 0 + CONTROL WORDS

| | |
|---|---|
| fris | hemd |
| kooi | riks |
| slim | nood |
| lijm | hars |
| mist | lomp |
| stil | vorm |
| boer | erin |
| krot | lans |
| loop | plek |
| bron | raad |

DUTCH TRANSLATION EQUIVALENTS - Overlap 1 + CONTROL WORDS

| | |
|---|---|
| fors | puin |
| poep | wenk |
| hert | zalm |
| kous | stal |
| kant | orde |
| buit | gade |
| geur | erin |
| maat | gids |
| paus | mouw |
| teug | drie |

DUTCH TRANSLATION EQUIVALENTS - Overlap 2 + CONTROL WORDS

| | |
|---|---|
| dood | taal |
| trom | keet |
| vlas | zulk |
| laan | doop |
| naam | kind |
| riet | heil |
| plak | Inkt |
| roet | smid |
| tijd | keer |
| waar | heen |

DUTCH TRANSLATION EQUIVALENTS - Overlap 3 + CONTROL WORDS

| | |
|---|---|
| beer | kalm |
| vlag | grof |
| goud | kast |
| haal | worp |
| kerk | ooit |
| merk | dief |
| melk | boel |

| | |
|---|---|
| peer | kruk |
| stap | kost |
| wiek | faam |

ENGLISH COGNATES + CONTROL WORDS

| | |
|---|---|
| band | cure |
| cape | flaw |
| chef | fuze |
| dame | gush |
| gong | lira |
| homo | pine |
| mama | skip |
| mini | snug |
| papa | tier |
| pass | view |
| pint | yawn |
| pose | bait |
| rank | bore |
| show | face |
| term | pack |
| tram | plum |
| vest | roam |
| wild | urge |
| wolf | veil |
| yoga | vide |

ENGLISH FALSE FRIENDS + CONTROL WORDS

| | |
|---|---|
| arts | bake |
| boot | chap |
| colt | dace |
| dank | eddy |
| gist | fang |
| hart | flea |
| last | wing |
| list | army |
| mare | beak |
| meet | fine |
| most | gaol |
| perk | glen |
| pond | hack |
| punt | hike |
| rein | hood |
| room | keep |
| slot | lash |
| trap | melt |
| veer | nick |
| wand | oboe |

ENGLISH TRANSLATION EQUIVALENTS - Overlap 0 + CONTROL WORDS

| | |
|---|---|
| airy | prim |
| cage | puff |
| cute | heed |
| glue | rift |
| haze | rump |
| hush | tidy |
| jack | bore |
| slum | axis |
| walk | draw |
| well | also |

ENGLISH TRANSLATION EQUIVALENTS - Overlap 1 + CONTROL WORDS

| | |
|---|---|
| bold | gene |
| crap | pout |
| deer | puff |
| hose | tier |
| lace | spit |
| loot | stab |
| odor | tidy |
| pall | weir |
| pope | gush |
| swig | vide |

ENGLISH TRANSLATION EQUIVALENTS - Overlap 2 + CONTROL WORDS

| | |
|---|---|
| dead | wife |
| drum | acre |
| flax | lira |
| lane | keen |
| name | fish |
| reed | knit |
| slab | pine |
| soot | pomp |
| time | back |
| ware | crux |

ENGLISH TRANSLATION EQUIVALENTS - Overlap 3 + CONTROL WORDS

| | |
|---|---|
| bear | soft |
| flag | luck |
| gold | drag |
| haul | visa |
| kirk | gene |
| mark | pour |
| milk | push |
| pear | rift |
| step | wear |
| wick | snug |