Radboud University Nijmegen

Faculty of Social Sciences

Expanding the Multilink lexicon with the French language

A collection of simulation studies

Thesis BSc Artificial Intelligence

Author:
Justin Greep

Supervisor:
prof. dr. Ton Dijkstra

Second reader:
dr. Frank Leoné

February 14, 2020

Abstract

Multilink is a computational model that simulates human word recognition and translation processes according to a localist-connectionist paradigm. The input word results in the activation of the target word, its orthographic neighbours, and semantically related words. In its current form, Multilink only accommodates bilingual Dutch-English word recognition and translation tasks. This thesis describes the extension of the current lexicon with French vocabulary. Furthermore, two simulation studies were performed to measure the performance of Multilink with its newly acquired lexicon.

The first study compared variations of the new French lexicon, such as using different frequency measures and the presence or absence of diacritical characters. These variations were tested by Multilink simulations of the lexical decision task. Simulation data (cycle times) were correlated with reaction times derived from the French Lexicon Project (Ferrand et al., 2010). A lexicon involving movie-derived frequencies produced a model-to-data correlation of 0.391 (n=832). The presence of diacritical characters did not produce a notable difference in correlation.

The second study simulated the empirical study by Peeters, Dijkstra, and Grainger (2013). This study examined the cognate facilitation effect by performing English-French lexical decision tasks. The results of this simulation produced a model-to-data correlation of 0.667 (n=166, $p < 0.0001$).

These results show that Multilink is capable of simulating word recognition processes involving the French language. Especially in the French-English bilingual domain, Multilink scores very well.

Contents

"Go back?" he thought. "No good at all! Go sideways? Impossible! Go forward? Only thing to do! On we go!"

— J.R.R. Tolkien, *The Hobbit, or There and Back Again*

1 Introduction

Mixing up words from foreign languages or simply forgetting how to speak a certain language when you are on holiday may seem familiar. As someone who speaks Dutch, English, and French quite well, it has happened to me a few times. I always get a laugh out of asking for 'three baguettes' instead of 'trois baguettes' accidentally. It is interesting to see that these errors happen especially with two foreign languages and do not concern the mother tongue.

There could be multiple reasons for why the errors occur between my second language (L2) and third language (L3). One could be that the phenomenon depends on the level of proficiency of each language. L2 could simply be more developed (stronger) than L3 and thus exert a bigger influence on speech. However, this raises the question why the errors do not involve the first language (L1), given that that is the most developed language. Alternatively, as the mother tongue, L1 might be 'marked' and entirely excluded from the phenomenon, suggesting that the battle of proficiency is only fought between foreign languages. Another account could be that these errors depend on the origin and similarity of the languages involved. Perhaps English and French are confused because they both have characteristics of Romance languages (although English also has Germanic properties). In this case, the same kind of mix-ups should occur, for instance, with Spanish as L2 and French as L3, but not with Spanish as L2 and German as L3. Or is the important point here that the L1 in this case is Dutch, a Germanic language?

The effect L1 has on L2 during lexical processing, or vice versa, can be simulated in several bilingual language processing models, like the BIA/BIA+ model and Multilink (these models will be discussed in a later chapter). However, no model is available as yet that can simulate the interaction between three or more languages at the same time. Luckily, Multilink can expanded with more languages due to its modular nature as a software programme. Thus, it can be turned into a new multilingual language processing model with more than two languages.

In fact, the goal of this thesis is to expand Multilink with the French language and test it for monolingual French and bilingual French-English conditions. This extension will show that the theoretical account and assumptions underlying Multilink can be generalised to a new language and trilinguals. In addition, it will allow Multilink to simulate a larger variety of monolingual and bilingual language processing tasks in terms of their response times and errors. We will test the performance of Multilink after its expansion by comparing the model simulations to empirical studies, correlating cycle times with reaction times (model-to-data comparison). Different model variants of Multilink will also be compared (model-to-model comparison).

In Chapter 2, we will set the stage for the Multilink extension and its simulations by reviewing the BIA, BIA+, and Multilink computational models. Chapter 3 will describe new simulations done with a monolingual French lexicon, as well as several model-to-model comparisons. In Chapter 4, a bilingual French-English simulation study is described and analysed. The final chapter of this thesis will discuss the findings, further research, and conclusions.

2 Computational Word Recognition Models

Implemented word recognition models allow for precise simulations of human word retrieval in various empirical studies. Such models can also help to predict what word recognition results may surface in future experiments. The models discussed in this chapter are primarily focused on simulating bilingual language processing.

We will discuss the available computational models in historical order, starting with the Bilingual Interactive Activation (BIA) model and its successor, BIA+, and then zooming in on Multilink itself. Structure, usage, and applications of these models, as well as differences and similarities will be examined.

2.1 The Bilingual Interactive Activation model

One of the first implemented bilingual models is the Bilingual Interactive Activation (BIA) model of visual word recognition (Dijkstra & Van Heuven, 1998). It is an extended version of the Interactive Activation model (McClelland & Rumelhart, 1981), which simulates the monolingual recognition of words using visual feature, letter, and word detectors.

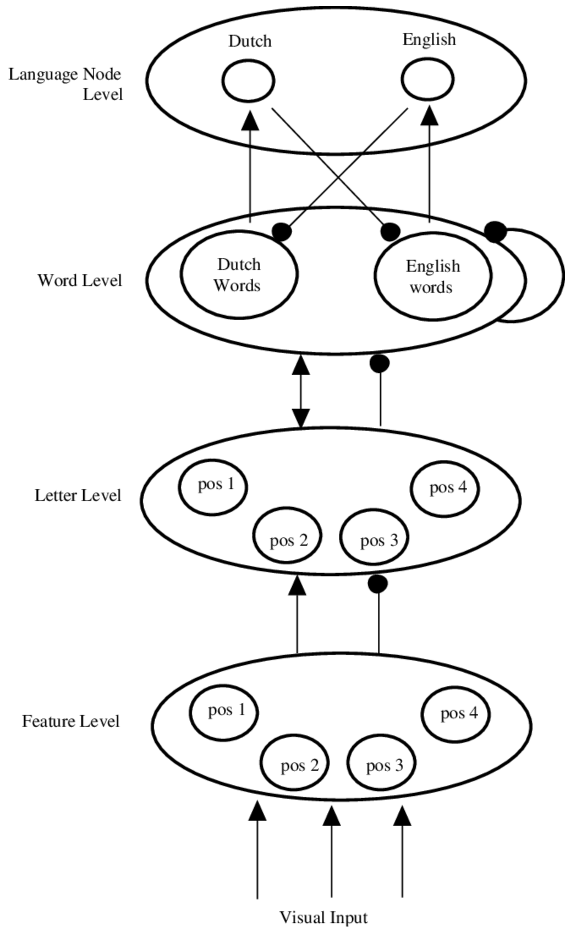


Figure 1: The BIA model for printed word recognition. Note: the arrows indicate an excitatory activation coming from the base of the arrow. The ball-headed lines indicate an inhibitory connection.

The BIA model simulates the same word recognition processing of words, but not only for monolinguals but also for bilinguals. It incorporates four layers, each consisting of nodes and connections between these nodes. The first layer represents the letter features. Each node in this level symbolises a letter feature, up to 14 nodes per letter position. The second layer has a total of 26 nodes per letter position and represents the letters. A lexicon of 1324 English and 978 Dutch four-letter and five-letter words is represented in the third layer and the nodes in the fourth layer mirror the English and Dutch languages. This fourth layer, along with the English-Dutch lexicon, is not present in the IA model. All nodes in each level are connected to each other, as well as to nodes of different levels. Figure 1 shows the basic architecture of the BIA model when a four-letter word is used as input.

When a string of letters that form a word or nonword is presented to the BIA model, all letter positions in that string are checked for the letter features in the first layer. The letter features that are present are then activated. Consequently, the letters that are made up of the activated letter features and their position are also activated. In turn, these activated letters activate their connected word nodes. In addition to sending activation to the corresponding language node, the active word nodes also send activation back to the letter layer, as well as inhibiting the activation of other word nodes. Lastly, an active language node also inhibits the words form the other language.

The word node that corresponds to the input string the most gains the highest activation and should cross a certain threshold. The time it takes to recognise a certain word is dependent on the characteristics of that word and how it is related to other words in the lexicon.

Considering that the BIA model is implemented, it can be used to simulate multiple bilingual language processing tasks, namely the generalised visual lexical decision task and the language decision task. By simulating these tasks, it is possible to come to a better understanding of how bilingual word processing transpires in the human brain.

However, the BIA model faces multiple limitations. Some features of bilingual word recognition are not sufficiently represented. Features such as phonology and semantics are not implemented in the model, and interlingual homographs and cognates are not specified. Furthermore, certain representational and functional aspects of the language nodes are unclear. These limitations were corrected in a revised model, called the BIA+ model. This model will be discussed in the next section.

2.2 The Bilingual Interactive Activation Plus model

The Bilingual Interactive Activation Plus (BIA+) model (Dijkstra & Van Heuven, 2002) can be seen as an upgraded BIA model. This revision was necessary, since new empirical data were made available at the time that disproved the choice of structure of the BIA model. The BIA+ model tries to improve these limitations in the BIA model.

To a large extent, the BIA+ model incorporates the BIA model, but with small alterations. The limitations of the BIA model described in the previous section, are largely solved in the BIA+ model. For example, (sub)lexical phonology, along with semantics, are now sufficiently represented in the model. Word recognition is now not only affected by cross-linguistic orthographic similarities, but also by similarities in cross-linguistic phonology and semantic resemblance.

As previously said, the BIA+ incorporates the BIA model. That is, the BIA+ model utilises a sublexical orthography layer similar to the letter and letter feature levels in the BIA model. Moreover, a lexical orthography layer like the word level in the BIA model is present, as well as similar language nodes.

A big difference with the BIA model is that the BIA+ model includes the lexical and sublexical phonology, and semantics. Additionally, a task control and decision system is connected to the word identification system. This task/decision control system stems from the Inhibitory Control (IC) model designed by (Green, 1998). The IC model explains that bilinguals can exert a certain amount of control on their processing in the lexico-semantic system. This is possible in different task conditions. The architecture of the BIA+ model, including its connection to the IC model, can be seen in figure 2.

The activation of nodes in this new system starts in the same manner as the BIA model and goes as follows: A letter string is given to the model. Several words that are orthographically similar to the string, are activated. The strength of the activation depends on the word's frequency. This results in a lower activation of L2 words, since they are often less frequent than L1 words. The orthographic words in turn activate the correct phonological and semantic representations.

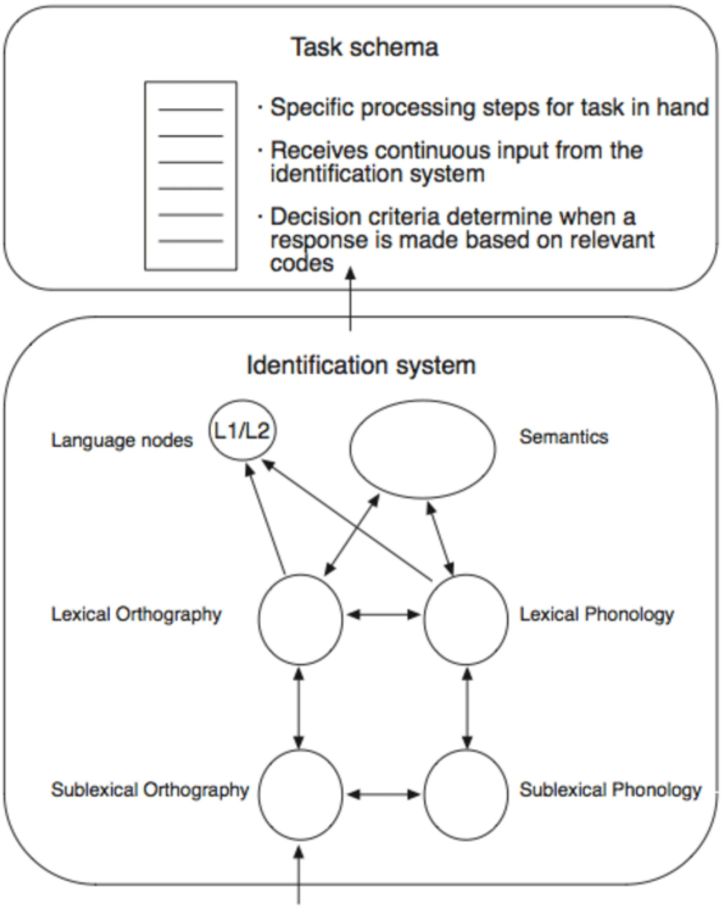


Figure 2: The BIA+ model architecture. The arrows describe the direction of activation between the layers. Inhibition within layers is not displayed.

Like in the BIA model, the BIA+ model uses two representations for interlingual homographs (words from different languages that share orthography, but not semantics). This was decided after evidence was found in multiple studies (Dijkstra, Van Jaarsveld, & Ten Brinke, 1998; Dijkstra, Grainger, & Van Heuven, 1999; Lemhöfer & Dijkstra, 2004), showing that each language contains a specific orthographic representation for interlingual homographs. No such evidence was found for cognates, so these were not implemented into BIA+.

Concerning the language nodes, a proposal is made that the languages nodes only represent the language membership of a word. This opposes the language nodes in the BIA model, which served as inhibitors for words from the opposite language. Besides, the paper indicates that the language information is found rather late in bilingual word recognition and the effect of language membership is thus minimal, since the word has often already been recognised before the language information is known.

Linguistic context effects describe the effect the context of a certain word has on its recognition. According to the BIA+ model, the recognition of words in a certain sentence context is influenced by the syntactic and semantic information of the context. Subsequently, BIA+ proposes that the activation of words in the Identification System is directly affected by this context information. Simultaneously, the model describes that the task/decision system is likely to be influenced by non-linguistic contexts, such as participant strategies.

Lastly, evidence was found that the model should indeed be divided by the Identification and Task/Decision system, even though the Identification system continuously sends information to the Task/Decision system. It is suggested that a word is recognised before a decision can be made. The opposite also makes an argument for this suggestion. A decision can be made before the world is recognised or when information is only partially available.

A later model that bares some resemblance to the BIA+ model, is Multilink. This model is the main focus of this paper, and will be discussed in the next section.

2.3 Multilink

Multilink is a computational language processing model that simulates human word recognition and translation (Dijkstra et al., 2019). It does this according to the mechanisms of a localist-connectionist paradigm.

In many respects, Multilink is similar to the BIA and BIA+ models. Like the BIA and BIA+ models, Multilink recognised words based mainly on the orthographic activation. However, Multilink is also capable of selecting words based on the semantic or phonological activation. Another difference between Multilink and its predecessor is that the older models can only handle the orthography of four- and five-letter words, while Multilink can handle words that contain between three and eight letters.

The input for the various cognitive tasks is always an orthographic representation of a word. The input word results in the activation of many lexical codes. Activation is passed through the model and resonates among codes. Semantically related words and orthographic neighbours may be coactivated over a set number of time cycles. For instance, when the cognitive task is to translate from English to Dutch, the input word DOG also leads to activation of words like BOG and LOG (orthographic neighbours), while CAT and PET are also activated (semantically related). Meanwhile, the Dutch translation for DOG, HOND, is also activated, as are the orthographic and semantic neighbours of HOND.

The Multilink model incorporates many thousands of nodes and connections between these nodes. These nodes represent the words in the lexicon, like DOG, HOND and CAT in the previous example. The connections all carry a value that indicates the facilitating or inhibiting effect the nodes have on each other. This means that the activation of one word can prevent another word from being activated, or at least lower the activation of that word. All words have a resting level activation (RLA), which is the level of word activation at time cycle 0. The RLA in the model depends on how often the word occurs per million words, and is set relative to the most frequent word in the training set.

Figure 3 shows the standard architecture of Multilink. The blue underscore illus-

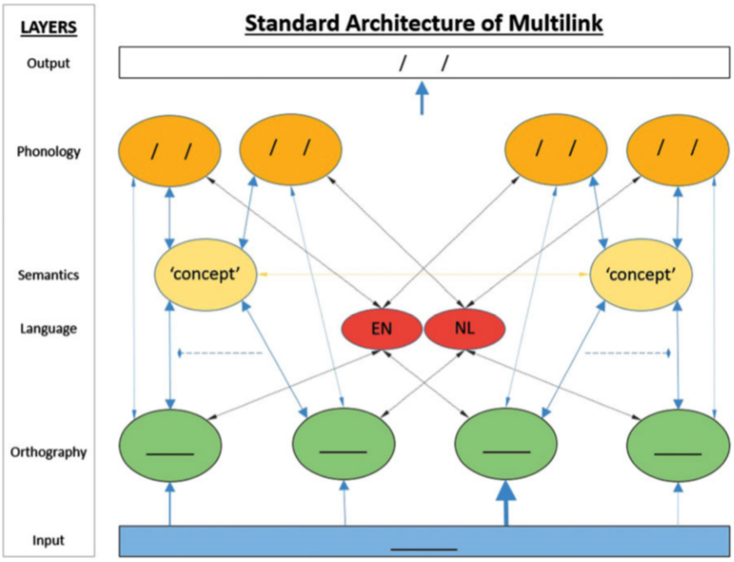


Figure 3: The standard architecture of Multilink

trates the input given by the user. The green underscores and yellow slashes indicate the orthographic and phonological nodes, respectively. The red nodes represent the current languages available in the lexicon, namely English (EN) and Dutch (NL). The white slashes on top of the image indicate a phonological output, however, since the output is task-dependent, it can be orthographic as well.

Currently, Multilink incorporates an integrated lexicon. This lexicon consists of 1466 Dutch words and their corresponding English translations. The words are present in both their orthographic and phonological form. The frequency of each words and its translations are accounted for as well. These frequencies are derived from the number of occurrences per million words in SUBTLEX-NL for Dutch words and in SUBTLEX-US for English words. The frequency for the orthographic and phonological representations are equal, since no good frequency database for phonology has been found yet, but the lexicon does feature An example of the lexicon can be found in Table 1.

Dutch O	Dutch P	Freq.	English O	English P	Freq.
CULTUUR	k}ltyr	11.62	CULTURE	kVlJ@R	3.485
CURSUS	k}rz}s	8.78	COURSE	k$s	121.805
DAAD	dat	16.65	DEED	did	2.3275
DAG	dAx	848.56	DAY	d1	200.455

Table 1: An example of the English-Dutch lexicon, where O and P mean 'Orthographic' and 'Phonological', respectively. Note: the phonology is written in X-SAMPA

Due to the localist-connectionist setup of its architecture, Multilink allows for a wide range of language process activities. For example, Multilink can work with words that differ in length, frequency and the amount of neighbours, all in a monolingual or bilingual way. It also allows for the usage of different word types in the lexicon, such as cognates (words that share orthography and semantics in multiple languages) or false friends (words that share orthography in multiple languages, but not semantics).

Additionally, the nodes inside Multilink can affect the activation of connected nodes.

This can result in either a facilitating or inhibiting effect. An example of the facilitating effect is the activation of orthographic neighbours in the lexical network. An example of the inhibitory effect is that activated neighbours may compete with each other, lowering one another's activation level.

Lastly, all of these activities can be invoked in several different language tasks, such as lexical decision, word translation, and language decision tasks. However, in this thesis paper, the focus will lie on the lexical decision task. The next chapter will discuss an empirical study, along with a series of simulations of this study, performed in Multilink.

3 Simulation of Monolingual Data

The goal of a computational model like Multilink is to provide a multidimensional architecture and processing framework that is isomorphic to the human word retrieval system. Empirical studies provide a relative good reflection of the 'real' processing system. If Multilink's simulation results compare well to the data gathered in an empirical study, this indicates that the model gives an adequate account of the human word processing involved. Thus, to measure the validity and performance involving the extended lexicon, the data of relevant empirical studies should be replicated by Multilink. To set the stage for our later simulations, we will first describe the empirical study used for those simulations. Afterwards, the simulations themselves will be described and their results will be examined.

3.1 French Lexicon Project

Megastudies are set up to collect huge amounts of data that other researchers can use for running large-scale analyses on word recognition and its variables. The French Lexicon Project (FLP) (Ferrand et al., 2010) is such a megastudy that gathered French lexical decision data to create a huge database. Other lexical data, such as word length, number of syllables, and frequencies were also collected and stored in this database. The FLP was setup to complement the English Lexicon Project (ELP) (Balota et al., 2007). The ELP is a lexical database like the FLP, but centred around the American English language. Projects like the FLP also exists for other languages, such as Chinese (Tse et al., 2017), Malay (Yap, Liow, Jalil, & Faizal, 2010) and Dutch (Keuleers, Diependaele, & Brysbaert, 2010).

In the study, participants where asked to perform a series of lexical decision tasks. In a lexical decision task, the participants are shown a string of letters that either formed a word or nonword (i.e., not a word in the target language). The participants have to press a button as quickly as possible if the letter string is a word, or another button if it is a nonword. Reaction times for 'yes' and 'no' responses are collected, as well as accuracy (error rates).

3.1.1 Participants

The participants consisted of 975 university students with a mean age of 21.4 years. All participants were native French speakers from two universities in France, Blaise Pascal University in Clermont-Ferrand and Paris Descartes University in Paris. The students had an average of 14.2 years of education.

3.1.2 Stimulus

A total of 38,840 words and 38,840 nonwords were shown in the lexical decision tasks. The words were gathered from Lexique 2 (New, Pallier, Brysbaert, & Ferrand, 2004) and Lexique 3 (New, Brysbaert, Veronis, & Pallier, 2007). Lexique is an online database that features a large collection of about 140,000 French words. These words are accompanied by various information, such as phonology, associated lemmas, number of syllables, grammatical category and frequencies of occurrence in different corpora. Lexique is freely available on the internet. (`http://www.lexique.org/`)

Initially, all words from Lexique with a frequency of 0.1 occurrences per million words (opm) or higher were selected, resulting in an initial list of 42,136 words. Names, fixed expressions and foreign words were filtered out of this list, as well as words that were unlikely to be known to the participants. Besides, words that had a low-frequent inflected form, e.g. plurals or feminine forms, were not removed from the list.

The nonword stimuli were custom made for this experiment. These nonwords were based on real words, but with slight alterations. Both monosyllabic and polysyllabic nonwords were used in the construction of nonwords to create monosyllabic and polysyllabic nonwords, respectively. The monosyllabic words were split up into rimes and onsets. Next, a huge collection of combinations between each individual rime and onset was formed. An example of these combinations can be seen in table 2. All real words and pseudo-homophones were removed from this collection. Lastly, a sample of this collection was taken such that it matched the real word selection in terms of mean number of neighbours and average string length.

Onset \ Rime	IEN	UIT
CH	CHIEN	CHUIT
N	NIEN	NUIT

Table 2: An example of onset and rime combinations for the words CHIEN (dog) and NUIT (night).

Polysyllabic nonwords were created by combining the individual syllables of real words. Once again, a sample was taken that matched the features of the real word selection.

3.1.3 Procedure

Each participant visited two different one-hour sessions. There was no more than one week in between the two sessions. During each session, the participants started with a practice round of 40 trials. Afterwards, they were shown a total of 500 words and 500 nonwords. Each trial started with two vertical lines in the centre of the screen with space between the two lines to clearly display the stimulus. These lines stayed on screen for 200 ms. Next, a stimulus was presented in the space between the vertical lines, which were still on screen. After the participant responded with one of the two buttons, the stimulus was erased. If the participant did not respond for 4 seconds, then the stimulus also disappeared. Each trial was concluded with a black screen for 1.5 seconds, after which the next trial started. The participants did not receive any feedback after the trials.

3.1.4 Results

Almost all participants had an average accuracy of at least 0.75 and an average RT below 1100 ms. The accuracy and mean RT of 21 participants were dropped, since they did not reach the same scores. All RT below 200 ms and above 2000 ms were identified as outliers. Furthermore, after calculating the standard deviation (SD) and mean for every participant, the RTs that deviated three or more SDs from the mean were considered outliers as well. Consequently, about 3.3% of all correct trials were rejected. Finally, the average RT for words was 730 ms (SD = 110) and 802 ms (SD = 120) for nonwords. The average percentage of errors was 8.9% (SD = 4.4) and 6.6% (SD = 3.9) for words and nonwords, respectively. There were about 25 observations per word.

As mentioned before, the main purpose of the FLP was to complement the ELP. However, there are some notable differences.

First of all, the FLP only gathered lexical decision times, whilst the ELP involved collecting both lexical decision data, as well as word naming data. In word naming trials, the participant is shown a word or nonword and then has to name the word as quick as possible.

Another difference between the two projects lies in the nonwords they used. In the ELP, the nonwords were made up by changing a single letter of a word, essentially creating a neighbour of that word. The FLP created nonwords that were similar to each other in orthography on the same level of similarity in orthography of words of the same length.

The presentation of the stimuli is also a notable difference. In the ELP, three asterisks were first shown, only to be followed by a blank interval screen. This screen was then replaced by the stimulus. Besides, the English stimuli were written in uppercase letters. The FLP used lowercase letters with the reason that participants were less used to seeing words written in uppercase letters.

Finally, the ELP gathered information for 40,481 English words, which is more than the amount of French words in the FLP. Furthermore, the ELP reached an average of 25 trials per word in the word naming tasks and 34 trials per word in the lexical decision tasks, which is notably higher than the FLP.

The ELP and the briefly mentioned Dutch Lexicon Project (DLP) have been used to measure Multilink's cycle time accuracy and to compare it to the BIA/BIA+ model. Dijkstra et al. (2019) described simulating lexical decision tasks in Multilink. The returned cycle times were then compared to the average RTs in the ELP and DLP, using Pearson's r correlations.

Simulations of the ELP and DLP returned satisfying results, so one could predict that simulations of the FLP will yield results equally as satisfying. Such simulations will be discussed in the next section.

3.2 French Simulations

This chapter describes multiple simulations that were run to test how Multilink performs with a monolingual French lexicon. Also, multiple changes to the lexicon were made, such as the removal of diacritical marks or a different source for the word frequencies was used. These variations were also tested to find the optimal configuration of the lexicon. The simulation outcomes are compared to the RTs extracted from the FLP.

3.2.1 Extending the Lexicon

In order to replicate French lexical decision tasks in Multilink, a French lexicon is needed. This lexicon was constructed by expanding the current lexicon.

First of all, a French translation for each entry in the lexicon was manually picked in order to ensure semantic equivalence with the Dutch and English translations. French words that were shorter than three or longer than eight characters were not used, and if no other translations would fit, the entire entry in the lexicon would not be used. The phonological representations of these newly found translations were extracted from Lexique 3.83 (New et al., 2004). These representations are written in X-SAMPA notation.

Besides the phonology, the frequency of the French translations was also extracted from Lexique. Lexique offers four different frequencies: The occurrence of the word per million words in books or in movies, and the occurrence of the lemma of the word per million words in books or in movies. The frequencies of the lemmas were not used in these simulations, since it does not correctly represent the used translations. Instead, the frequencies derived from movies were used in the new lexicon, but a variation with the book frequencies will be discussed later in this chapter.

The expansion resulted in a trilingual Dutch-English-French lexicon. A monolingual French lexicon was created by isolating the French section of the trilingual lexicon. This monolingual lexicon served as a basis for the first set of simulations. It contained 1270 French nouns and adjectives. All words are written in uppercase letters.

3.2.2 Variation on the Lexicon: Frequency Sets

Since Lexique offers two sets of frequencies, it would be interesting to see if their would be any difference between the two in Multilink. The difference in these sets lies in their sources. One set contains frequencies derived from movie subtitles, the other contains frequencies that are derived from books.

To find the set that fits the new lexicon best, two sets of simulations had to be run. Both sets simulated lexical decision tasks. The outcome of these simulations was correlated to their corresponding RTs extracted from the FLP. The difference between the two simulation sets rested in the content of the lexicon. One set of simulations (Set_b) used the frequencies derived from books, while the other set (Set_f) used the frequencies derived from movie subtitles. The orthography and phonology remained the same in both lexicons.

Both sets were run with the entire lexicon as input, so a total of 1270 trials were simulated in Multilink with each lexicon. A certain amount of cycle times was returned if the input word was recognised. The trials that were not recognised by Multilink before the final cycle (40) had been reached, were excluded from the analyses. The rest of the cycle times were correlated to the RTs from the FLP.

3.2.3 Results

As a result, there was a correlation of 0.313 (n=837) for Set_b and a correlation of 0.391 (n=832) for Set_m. So, it turns out that when a lexicon with book frequencies is used in Multilink simulations, it correlates slightly lower with RTs than when a lexicon is used that contains movie frequencies. The data can be seen in figure 4.

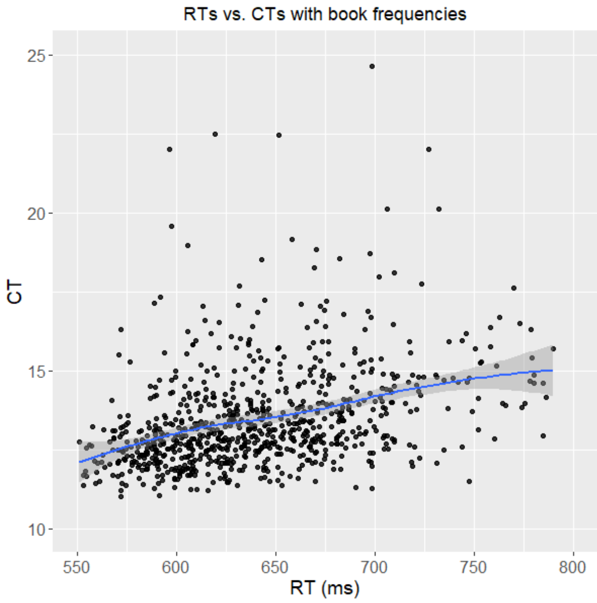
We see that in both frequency cases, that there is not really a good correlation between the FLP reaction times and the Multilink generated cycle times. In the graphs, it becomes clear that there exists some noise in the data. Namely, the upper part of the graphs show quite a lot of potential outliers. Due to limited time, these possible outliers have not been isolated and analysed, but it may explain the relatively low correlations.

The two frequency sets do correlate relatively high with each other, namely $\rho = 0.848, n = 87406$. However, there could be some reasons for the difference in the correlations between the RTs and CTs. For example, books are often more descriptive in nature, movies, and especially movie subtitles, are primarily focused on conversations. Certain words that are more often used to describe scenes, landscapes or expressions are likely to be more frequent in books than in movies. Subsequently, words that appear more often in conversation are much more likely to have higher frequency in movie subtitles. For example, the frequency of CHÉRIE has a frequency of 98.95 opm in the movie frequency set, but a frequency of 17.64 opm in the book frequency set.

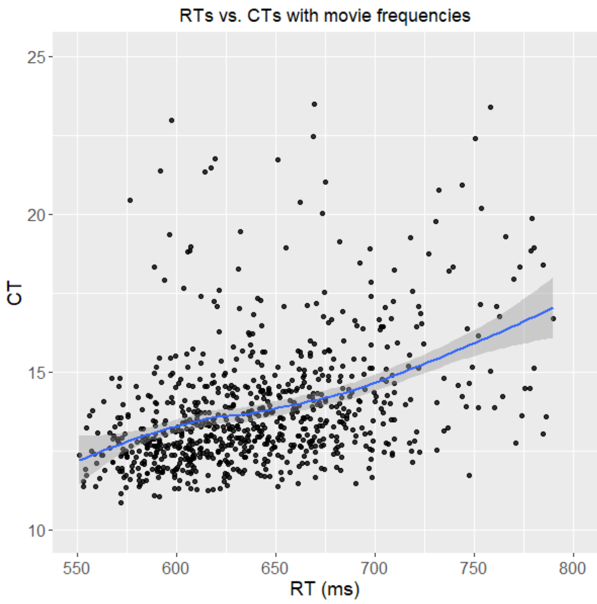
It is interesting to see that the highest correlation is only 0.391, since simulations that were run with Multilink to replicate the ELP and DLP data, correlated much higher with the real world data than this simulation did. A correlation of 0.630 and 0.514 were achieved with the DLP and ELP, respectively.

Consequently, A reason for the difference could be that the words in the ELP and DLP had much more observations, namely 34 and 39 on average, respectively. The words in the FLP, however, had only been observed for about 24 times. So the RTs in the FLP could reflect the real world less than the ELP and DLP. It could also be that Multilink is able to run monolingual simulations, but just not as well for the French language.

At last, one of the big differences between the French language and the Dutch and English languages, is the multitude of diacritical characters. These characters could of course also be a reason for the decrease in correlation. The next section will look into the effect of these diacritical characters on the performance of Multilink.



(a)



(b)

Figure 4: (a) displays the Multilink CTs generated with book-derived frequencies, whereas (b) shows the CTs generated with movie-derived frequencies. In both graphs, the CTs are plotted against the FLP reaction times.

3.2.4 Variation on the Lexicon: Diacritical Characters

French words are characterised by the large amount of diacritical characters. Marks, such as the acute accent (é), grave accent (è) or the cedilla (ç) are very common in French. Often these characters change the phonology of a word. Take for example ENTRE (between) and ENTRÉ (past particle of ENTRER, or to enter). The last E in ENTRE has an open-mid front rounded vowel pronunciation, like the i in 'bird'. The É in ENTRÉ is an close-mid front unrounded vowel, as ay in 'may'.

Because the affect the characteristics of the printed input signal, the diacritical marks can make a difference in word recognition. For instance, the number of similar words may be dependent on whether the diacritical marks are considered or not. This is probably also the case for Multilink, but has not been tested: The Dutch-English Multilink lexicon does not contain any Dutch or English words with diacritical characters.

In general, English words contain almost no diacritical marks. Most words that do, originate from French, Spanish, or German. The same holds for Dutch, although, the diaeresis is a lot more prominent than other diacritical marks.

In Multilink, about 17% of the words in the monolingual French lexicon contain one or more diacritical characters . To test whether diacritical characters make a difference in Multilink's monolingual French lexical decision performance, two series of simulations were run.

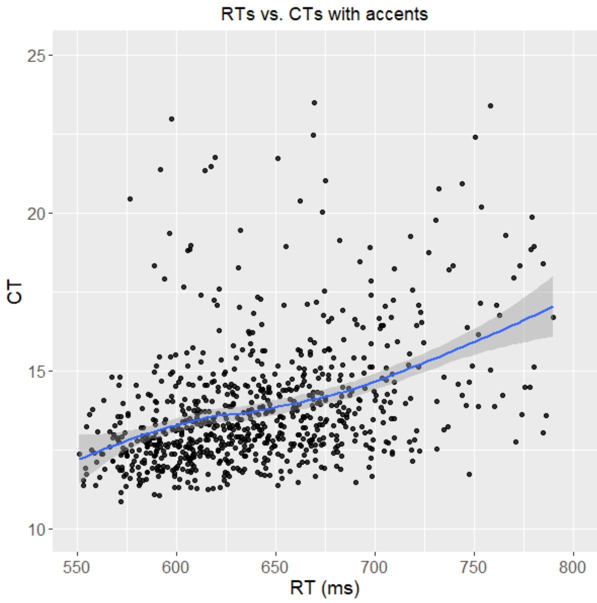
The first set was run with the normal lexicon, so with diacritical characters, and the movie-derived frequencies, since they scored the highest correlation in the last round of simulations. The input contained all the words in the lexicon, so each simulation was run on a different word from the lexicon.

The second set used a variation on the normal lexicon. Namely, all letters that contained a diacritical mark were replaced the letter without the mark, e.g.: É was replaced by E. Obviously, the words in the input underwent the same transformation. In the end, the lexicon and input did not contain any diacritical marks. The CTs generated by Multilink were once again compared to the RTs in the FLP using Pearson's r correlations.

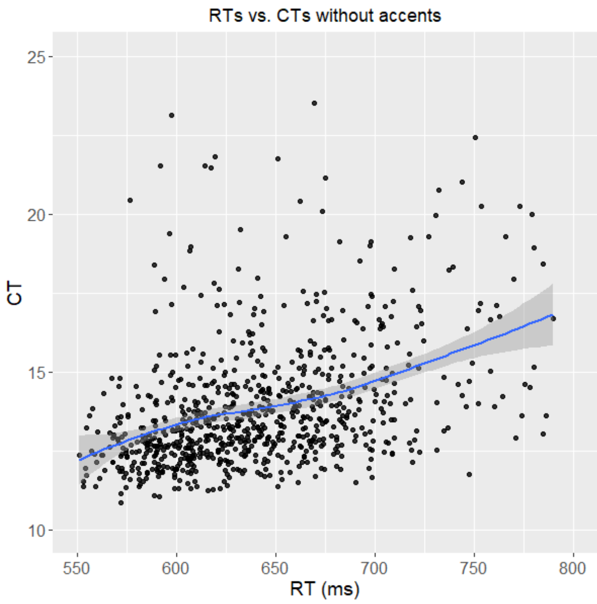
3.2.5 Results

A correlation of 0.388 (n=834) was found between the CTs generated by Multilink without diacritical characters, and the FLP RTs. As we have seen already in the previous section, a correlation of 0.391 (n=832) was found between the CTs generated by Multilink, using the normal lexicon, and the RTs gathered from the FLP. The data can be seen figure 5.

Apparently, using no diacritical characters in the lexicon and input makes almost no difference for the correlation between the CTs and the FLP RTs. This can also be seen in the graphs, since they are almost identical. The simulations performed without accents correlates slightly less with the real world data, but the difference seems to be negligible.



(a)



(b)

Figure 5: (a) shows Multilink CTs generated with a lexicon that contains diacritical characters, while (b) shows CTs generated with a lexicon without diacritical characters. The CTs are plotted against the FLP reaction times in both graphs.

4 Simulation of Bilingual Data

The performance of Multilink of simulating monolingual lexical decision tasks can be easily assessed against empirical data in these huge databases. Unfortunately, no such databases exists for bilingual lexical decision data. To check Multilink's bilingual word recognition performance, we can compare the cycle times to RTs gathered in much smaller empirical studies.

This chapter will first describe such a smaller empirical study that gathered English-French lexical decision data. Afterwards, the simulation of that empirical study will be described and the chapter will conclude with the results of that simulation.

4.1 French-English Lexical Decision Study

A study done by Peeters, Dijkstra, and Grainger (2013) involved analysing how late French-English bilinguals process identical cognates (words that have the same form and meaning in multiple languages, e.g. FRUIT). This analysis is done by comparing RTs and Event-Related Potentials (ERP) with identical cognates and control words.

4.1.1 Participants

Nineteen participants with an average age of 22.3 years old engaged in both the RT and ERP study. All participants were native French students and had English as their L2. The average age in which they started learning English was 10. Before the experiments, the participants assessed their own French and English level of proficiency via a questionnaire. Overall, on a Likert scale from 1 to 7 (1 = unable, 7 = expert), the mean level of proficiency ranked on 6.9 and 5.5 for French and English, respectively. Also, the average daily usage for French was 75% and for English 25%.

4.1.2 Stimulus

The stimuli for the combined RT and ERP experiment consisted of 120 English-French identical cognates, 120 English control words and 240 nonwords that resembled real English words, making a total of 480 trials. The words were collected from the CELEX database (Baayen, Piepenbrock, & Van Rijn, 1993), and all of them are most often encountered as noun. Additionally, all stimulus lengths were between the 3 and 9 characters.

A distinction was made between four types of cognates based on the relative frequency each cognate has in both the English and French language. They can be cognates with a high frequency in both languages (HEHF), high frequency in English, but a low frequency in French (HELF) or vice versa (LEHF), or they can have a low frequency in both languages (LELF). However, what complicated the matter is that the frequency of identical cognates in the two languages involved is usually highly correlated (Schepens, Dijkstra, & Grootjen, 2012).

The four types of cognates equally distributed the total amount of cognates, so thirty words per cognate type. The orthography of these cognates was equal in both French and English. The non-cognate control words could be split up into two types: Highly frequent (HE) and infrequent (LE) English words. All cognates and control word had a lemma frequency of at least one occurrence per million in CELEX. The nonwords were constructed by changing one letter in nouns that were not used as a cognate or control word, using Wuggy (Keuleers & Brysbaert, 2010).

All the stimuli were combined in randomised lists and divided into four blocks. No blocks were identical, and were constructed in such a way that no more than three words or nonwords would appear in a row. Furthermore, cognates from the same type would not succeed each other more than once. Table 3 shows some examples of the different stimuli.

Stimulus	Example
HEHF	ACCENT
HELF	CAMPUS
LEHF	REVOLVER
LELF	INTRUSION
HE	ENTRANCE
LE	SQUIRREL
NONWORD	CUDBOADS

Table 3: Examples of the stimuli used in the lexical decision study

4.1.3 Procedure

At the start of the experiments, the participants were instructed to make lexical decisions as quickly as possible. They were also asked to not blink unless a certain symbol would appear on the screen. The stimuli were presented on a black screen with white letters. Before each stimulus, a cross was displayed on the screen to indicate a fixation point. After 200 ms, this fixation cross disappeared and was followed by a stimulus. This stimulus remained on the screen 1300 ms, after which the stimulus disappeared and the blink symbol appeared. This symbol was displayed for 2500 ms and succeeded by a blank, black screen. The next trial would start after 300 ms. The participants were allowed to response for up to 1500 ms, starting at the stimulus onset. Decisions that were made after this deadline were considered as errors. After each block of stimuli, the participants could take a break for as long as they prefered.

The ERP data were gathered with an electroencephalogram (EEG) that was recording continuously. 64 Ag-AgCl electrodes were fixed to an elastic cap, which was placed according to the 10-10 international system. Additionally, two electrodes were placed below the left eye and on the lateral canthus to the right of the right eye to record blinks or vertical eye movements, and horizontal eye movements or saccades. Also, an electrode was placed on the left and right mastoid bone.

4.1.4 Results

About 10.84% of the collected data was an error and thus removed. All RTs that deviated more than 2.5 times the SD from the participant's mean were considered outliers. Variance analyses were performed on average RTs and error rate on an subject level (F_1) and on a item level (F_2). The average RT of recognition of words was 705 ms, and of nonwords 812 ms. Overall, the identical cognates (694 ms) scored a significantly faster RT than the control words (726 ms). Furthermore, HE words (669 ms) were recognised significantly faster than LE words (752 ms). Also, the participants responded significantly faster to the cognates with a high English frequency than to cognates with a low English frequency. The same held for French: Highly frequent cognates in French were responded to significantly faster than to infrequent cognates.

On average, the error rate was 10.02% for words and 11.67% for nonwords. The average error rate in control words was 11.01% and 9.04% in cognates. There was no significant difference in error rate in cognate recognition and control word recognition. However, there was a significantly lower error rate in HE recognition than in LE recognition. Additionally, there were fewer errors made in recognition of HEHF and HELF than in LEHF and LELF. Again, the same holds for the French language: fewer errors were made in recognition of HEHF and LEHF compared to HELF and LELF.

Overall, an N400 cognate facilitation effect in the parietal areas was visible in the ERP data, as well as distributed effect of the English frequency in the N400 time window. It was also shown that cognates produced a late positivity (600-900ms) when compared to the control words. Additionally, by comparing the ERP data for the four types of cognates, the English frequency produced a long and wide-spread N400 effect, while the French frequency produced a shorter, anterior N400 effect. And finally, a significantly more negative N400 was found with LEHF cognates than with LELF cognates.

To conclude, this study shows that there does exists a cognate facilitation effect in bilinguals, i.e. cognates that exist in both L1 and L2 are recognised faster and with less errors than non-cognates. However, there was no evidence that cognates that have a high frequency in L1 and L2 (HEHF) are recognised faster than cognates that have a high frequency in L1 but a low frequency in L2 (LEHF).

Additionally, the authors proposed a certain two-morpheme view accounted best for the behavioural and electrophysiological data. According to this two-morpheme view, cognates are represented by one morphological and one phonological form for each language in the human mind. These form also come with there own characteristics that are suitable for each specific language, such as frequency, gender or plural form.

The next section will walk through the procedure and results of the simulation of the study described above.

4.2 French-English Simulation

From the previous chapter, it can be deduced that the French part of the Multilink lexicon should use the movie-derived frequency, because the monolingual lexical decision task simulations correlated higher with the real world data when these frequencies were applied in the lexicon. The alternative, book-derived frequencies, did not perform as well.

These frequencies will now also be used in the bilingual English-French lexicon, and possible in the future trilingual Dutch-English-French lexicon. As for diacritical characters, they will remain in the lexicon, since the effect of these character on Multilink's word recognition is negligible. Secondly, the characters represent the French language better than when they would be removed from the lexicon. Thirdly, since no accent removal is needed, it could be easier in the future the expand the French lexicon with more content.

This chapter will discuss the simulation of the study done by Peeters, Dijkstra, and Grainger (2013). This study was earlier described in section 4. To quickly summarise, the study describe a series of lexical decision tasks performed by French-English bilinguals. The ERPs and RTs were recorded during the experiment to analyse the cognates facilitation effect. The study concluded that there is indeed such an effect.

For this simulation study, no ERP data will be simulated or analysed. The lexical decision tasks, however, will be simulated in Multilink. These simulations will be run with the bilingual English-French lexicon as described above. The input for the simulation series will be the same as the input from the replicated study, minus the nonwords. Consequently, 240 lexical decision tasks will be performed by Multilink: 120 on English-French cognates and 120 on words that only occur in the English language.

The cycle times produced by Multilink will be compared with the reaction times collected by Peeters et al. (2013) using Pearson's r correlation.

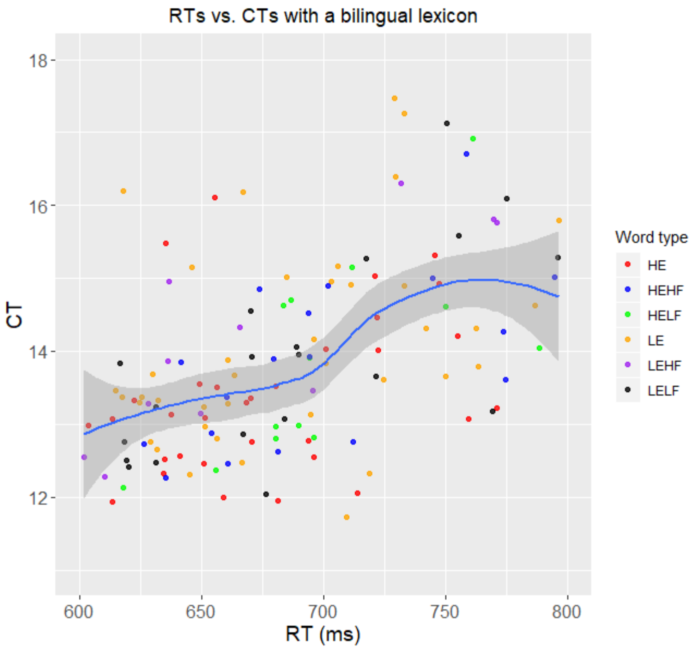
4.3 Results

A total of 166 words out of 240 words were correctly recognised by Multilink. Almost all words that were not recognised, were not present in the lexicon. These words were not taken into account in the analyses. Comparing the generated CTs with the RTs, a correlation of 0.667 ($p < 0.005$) was found. The RTs and CTs plotted against the word frequencies can be found in figure 6.

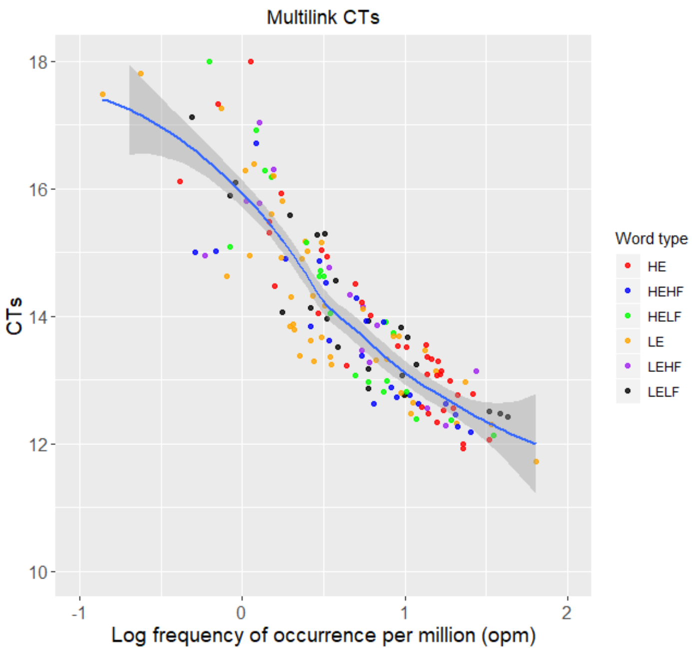
The obtained correlation is quite high, especially compared to the previous simulation study. This high correlation is also visible in the graphs in figure 6. It appears that the the Multilink CTs correlate really well with the real word RTs. It is clearly shown that the words and cognates that are frequent in at least one language, are recognised quickly by Multilink. The words that are less frequent in English, are recognised relatively late, which corresponds well with the real world.

Subsequently, the high correlation shows that Multilink performs quite well with a English-French lexicon when simulating lexical decision tasks. However, the error rate is quite high as well. 69% of the stimuli have been recognised. The high error rate could simply be explained by the lack of these stimuli in the lexicon. A word such as CORRIDOR is not present in the lexicon, but it is highly likely that an English speaker has encountered this word at least once.

Besides, the number of items that the comparisons in the previous chapter used, was more than five times higher than the sample size in this study. So increasing the sample size in this study, by introducing more input words or extending the lexicon content-wise, could affect the current correlation. Nonetheless, the obtained correlation is a commendable starting point, for this is the first time a French-English bilingual lexicon is used in Multilink.



(a)



(b)

Figure 6: The results of the second simulation study. Graph (a) displays the Multilink CTs plotted against the RTs gathered by Peeters et al. (2013). (b) shows the CTs plotted against the log frequencies.

5 General Discussion

In the present study, Multilink was extended to include a French lexicon. In addition, simulation studies were done to assess the performance of the new version of Multilink. Chapter 2 described multiple computational word recognition models, namely BIA, BIA+, and Multilink. Chapters 3 and 4 described two simulation studies.

The first simulation study investigated the processing consequences of different frequency measures and the presence or absence of diacritical characters. This study was divided into to two parts. In the first part of the study, two sets of frequencies were gathered from Lexique 3.83, in particular concerning movies and books. Simulations of the French lexical decision task with both types of frequencies in Multilink were correlated with reaction times in the French Lexicon Project. The resulting two correlations can be found in the top row of table 4 (Accents).

	Movies	Books
Accents	0.391 (832)	0.313 (837)
No Accents	0.388 (834)	0.303 (826)

Table 4: The correlations obtained for the first simulation study. Following the correlation, the number of trials is given in parentheses. Note: The bottom right correlation was calculated for completeness, and was not discussed in this paper.

The second part of the first simulation study also considered the effect of diacritical characters, such as É and Ç. Once again, two variations of the lexicon were used in simulations of the lexical decision task: one variation with these diacritical characters and one without. The cycle times produced by Multilink in these simulations were also correlated to the RTs in the FLP, and can be found in the first column of table 4 (Movies).

The second simulation study was concerned with an English-French experiment by Peeters, Dijkstra, and Grainger (2013). This study looked into the cognate facilitation effect in the lexical decision performance of French-English bilinguals. It produced English lexical decision data for cognates and non-cognates that were high or low in frequency in either language. Multilink simulated the English lexical decision task with an French-English lexicon. Comparing the output of the simulations with the RTs from the study resulted in an impressive correlation of 0.667 (n=166).

These simulation studies indicate that Multilink performs quite well with an bilingual English-French lexicon while performing lexical decision tasks. However, the simulation of monolingual French words worked less well.

Nevertheless, the large difference in results in the simulations for FLP, ELP, and DLP (seen in table 5) begs for an explanation. This difference could be explained in terms of the data that were used to correlate the cycle times with. As shown in figure 4, there is a lot of noise in the RTs from the FLP. The graphs of RTs in the paper by Dijkstra et al. (2019) show a lot less noise and generally more compact data. Table 5 also shows the average observations per item in the experiments of the lexical projects. The FLP has fewer observations per item, which most likely have contributed to the amount of noise in the data.

Another reason for the disparities in correlation could be sought in the mapping between orthography and phonology (O-P) for French. Dutch has a 'shallow orthography' and a very straightforward O-P mapping. There are not many words that share orthography but differ in pronunciation, or words that share pronunciation but differ in orthography. French, however, does feature these types of words. For example, VIN, VINGT and VAIN share their phonology, but differ in orthography. Meanwhile, English

	DLP	ELP	FLP
ρ	0.630	0.514	0.391
n	214	227	832
avg. obs. per item	39	34	25

Table 5: Correlations of Multilink CTs and lexicon project RTs, the amount of trials for each correlation and the average observations per item in the lexicon projects

	NL-E	F-E
ρ	0.687	0.667
n	204	166

Table 6: The correlations gathered in Dutch-English and French-English simulations studies.

features these word types as well, such as WHERE, WEAR and WARE. Additionally, it also contains a lot of words that (partially) share orthography, but differ in phonology, for example THROUGH, TOUGH, and THOUGH. These intertwined O-P mappings could complicate cycle or reaction times. However, that would still raise the question why the ELP simulation did result in a higher correlation than the FLP simulation.

Table 6 shows the correlations gathered in bilingual simulation studies. As can be seen, the correlation of the Dutch-English simulation study is slightly lower than the correlation of the French-English simulation study. However, that study was performed with a lot more trials. Nonetheless, the correlations are quite close to each other, and thus both Dutch-English simulations, as well as French-English simulations, perform equally as good.

The outcomes of the bilingual simulation study clearly deviate from those of the monolingual simulation study. The difference in performance could be explained by the difference in sample size. Most FLP-replicating simulations performed around 832 trials. This is a big contrast to the 166 trials run in the bilingual situation. The smaller number of trials in the bilingual simulations could be attributed to small amount of available English-French bilingual data on the internet. Unfortunately, huge databases like the FLP do not exists for bilingual lexical decision data.

Furthermore, the frequencies in the lexicon do not accurately represent language proficiency in humans. In the study by Peeters et al., the participants used French three times as much as English and personally rated themselves as more proficient in French than English. If this could be resembled in the lexicon, then the simulations may compare even better to the real world data.

5.1 Further Research

Now that a French lexicon has been added to Multilink, this opens up a ton of new simulation possibilities. Not only French monolingual and English-French bilingual simulations can now be performed, but also Dutch-French bilingual, or even Dutch-English-French trilingual tasks can be addressed.

Such trilingual lexical decision simulations were considered for this thesis paper, but considering that trilingual Dutch-English-French lexical decision data were not readily available, the simulations were not performed. Gathering such data was not viable in the allotted time frame and falls outside of the scope of this paper. Collecting such data and comparing them to Multilink simulations could be the topic of a future research paper. For example, the data gathered by a study that focused on cognates and interlingual homographs processing in trilinguals (Biloushchenko, 2017) may be simulated in Multilink in the near future.

Additionally, it would be interesting to see how well word translation is conducted in Multilink, now that the French language can be utilised.This allows a test of how language distance (e.g., Dutch-German vs Dutch-French) and language family (Germanic vs Romance) affect performance in multilinguals. Also interesting is the comparison of translation performance for Dutch-French cognates and non-cognates in order to study word translation in trilinguals.

Expanding the lexicon with respect to the number of words may yield different results as well. Especially in the bilingual simulation study, many words were not recognised by Multilink. Almost all of these words did not occur in the lexicon. Adding these words to the lexicon and repeating the simulation study might lead to a different correlation, since the sample size for the correlation would be higher.

Finally, a future paper could extend the Multilink lexicon with still other languages, preferably those that share their origin with French, like Spanish or Italian. A Spanish lexicon has already been made and RTs gathered from SPALEX (Aguasvivas et al., 2018) were simulated. This resulted in a correlation of 0.643 (n=1456). This correlation is in line with the earlier ELP and DLP simulations, so it reinforces the idea that the FLP data are wrong.

Yet another future paper could investigate what facilitating effects languages of the same origin have on each other in Multilink. For example, a comparison can be made between French-Spanish simulations and French-English simulations to see in which case Multilink performs better. The question whether the origin of the language matter, plays a valuable role here.

5.2 Conclusions

The goal of this paper was to expand the Multilink lexicon with French vocabulary. In the first chapter, the word recognition models BIA/BIA+ and Multilink were discussed. The latter served as the main focus for the current paper. Additionally, two empirical studies were discussed. The French Lexicon Project involved the creation of a huge lexical decision database for the French language. The study by Peeters, Dijkstra, and Grainger (2013) looked into the cognate facilitation effect in French-English bilinguals.

Next, two simulations studies were done to test the extended version of Multilink. In the first, the FLP was used to measure the performance of lexical decision simulations by Multilink for three different configurations of the monolingual lexicon. Two configurations were distinguished by frequency sets, and the other two were differentiated by the presence of diacritical characters. Simulations run by Multilink with a lexicon using movie-derived frequencies and diacritical characters scored the highest correlation score, namely 0.391 (n=832). It is likely that the data contributed to the low correlation.

A second simulation study tested Multilink's performance in French-English bilingual simulations. The study by Peeters et al. (2013) was replicated with Multilink while using a French-English lexicon. A comparison between the study and simulation results produced an impressive correlation of 0.667 (n=166). Simultaneously, this correlation marks the successful completion of this paper's research goal.

This paper shows that Multilink is indeed capable of handling new languages in the lexicon, especially when it has to simulate bilinguals. By extending the Multilink lexicon, we have gotten one step closer to truly simulating human word retrieval. Additionally, the high correlation in the bilingual simulation study shows a promising future for Multilink. One that can give us an even better view of human word recognition. And maybe, just maybe, one that will finally tell me why I always order 'three baguettes' in a French bakery.

References

Aguasvivas, J. A., Carreiras, M., Brysbaert, M., Mandera, P., Keuleers, E., & Duñabeitia, J. A. (2018). Spalex: A spanish lexical decision database from a massive online data collection. *Frontiers in psychology*, *9*, 2156.

Baayen, R., Piepenbrock, R., & Van Rijn, H. (1993). The celex lexical database (cd-rom) philadelphia. *PA: Linguistic Data Consortium, University of Pennsylvania*.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... Treiman, R. (2007). The english lexicon project. *Behavior research methods*, *39*(3), 445–459.

Biloushchenko, I. B. (2017). *How trilinguals process cognates and interlingual homographs: the effects of activation, decision, and cognitive control* (Unpublished doctoral dissertation). University of Antwerp.

Dijkstra, T., Grainger, J., & Van Heuven, W. J. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and language*, *41*(4), 496–518.

Dijkstra, T., & Van Heuven, W. (1998). The bia-model and bilingual word recognition. *Localist Connectionist Approaches to Human Cognition*, 189–225.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, *5*(3), 175–197.

Dijkstra, T., Van Jaarsveld, H., & Ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition*, *1*(1), 51–66.

Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink: a computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, *22*(4), 657–679.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., ... Pallier, C. (2010). The french lexicon project: Lexical decision data for 38,840 french words and 38,840 pseudowords. *Behavior Research Methods*, *42*(2), 488–496.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition*, *1*(2), 67–81.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior research methods*, *42*(3), 627–633.

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*, 174. Retrieved from `https://www.frontiersin.org/article/10.3389/fpsyg.2010.00174` doi: 10.3389/fpsyg.2010.00174

Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition*, *32*(4), 533–550.

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.

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied psycholinguistics*, *28*(4), 661–677.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new french lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 516–524.

Peeters, D., Dijkstra, T., & Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: Rt and erp effects. *Journal of Memory and*

Language, *68*(4), 315–332.

Schepens, J., Dijkstra, T., & Grootjen, F. (2012). Distributions of cognates in europe as based on levenshtein distance. *Bilingualism: Language and Cognition*, *15*(1), 157–166.

Tse, C.-S., Yap, M. J., Chan, Y.-L., Sze, W. P., Shaoul, C., & Lin, D. (2017). The chinese lexicon project: A megastudy of lexical decision performance for 25,000+ traditional chinese two-character compound words. *Behavior Research Methods*, *49*(4), 1503–1519.

Yap, M. J., Liow, S. J. R., Jalil, S. B., & Faizal, S. S. B. (2010). The malay lexicon project: A database of lexical statistics for 9,592 words. *Behavior Research Methods*, *42*(4), 992–1003.