

**Radboud University Nijmegen**



MSC THESIS IN ARTIFICIAL INTELLIGENCE

---

# **Multilink: a computational model for word translation**

---

*Author:*

Erik Lormans

eriklormans@hotmail.com

*Supervisors:*

Prof. Dr. Ton Dijkstra

Dr. Ida Sprinkhuizen-Kuyper

May 25, 2012

All bilinguals, even children, are capable of quickly translating words from one language into another. This is surprising, because word translation requires complex interactions of language comprehension and production under executive control. This thesis describes an implemented localist-connectionist model, called Multilink, that describes the word translation process in the way humans perform the task. It accounts for the performance of high and low second language (L2) proficient bilinguals in tasks like forward ( $L1 \rightarrow L2$ ) and backward ( $L2 \rightarrow L1$ ) word translation, as well as lexical decision and language decision. Simulations are presented (based on a large set of stimuli; 4017 English words and 2180 Dutch words) that compare Multilink to monolingual and bilingual Interactive Activation models and consider the recognition of cognates (form-similar translation equivalents) and the translation of non-cognates.

In the monolingual simulations remarkably high correlations were found between Multilink and empirical data. A model-to-model comparison of Multilink to the Interactive Activation model yielded a favorable outcome for Multilink. Only on the word length effect, the model did not fully meet our expectations. In the bilingual domain, a model-to-model comparison between Multilink and the Bilingual Interactive Activation model again yielded a favorable outcome for Multilink. The word length effect was also re-examined in the bilingual domain, this time with outcomes that were in line with our expectations. Simulations on cognates and non-cognates processing show that Multilink is capable of handling words of cross-linguistic similarities and differences. Finally, in the translation simulations of non-cognates, the same asymmetry was found as predicted by the Revised Hierarchical Model, a theoretical model of word translation. However, where the Revised Hierarchical Model assumes this asymmetry is the result of word form links between L1 and L2 word form representations, Multilink observed the same asymmetry in forward and backward translation without the presence of word form links between L1 and L2 word form representations. Thus, the model shows that the empirical data do not require the presence of word form links between L1 and L2 word form representations as suggested by the Revised Hierarchical Model.

---

Special thanks to Ton Dijkstra and Ida Sprinkhuizen-Kuyper for their guidance and valuable advice. Thanks also to Steven Rekké for his preceding work and assistance in the translation simulations. And, last but not least, to Job Schepens for providing the Dutch translations.

## Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	The complexity of word translation . . . . .	4
1.2	Models of bilingual word retrieval . . . . .	7
1.3	Empirical studies on word translation . . . . .	12
1.3.1	Translation recognition . . . . .	12
1.3.2	Translation production . . . . .	14
1.4	Multilink . . . . .	17
1.4.1	Activation functions and parameter settings . . . . .	20
1.4.2	The modified Levenshtein distance . . . . .	21
<b>2</b>	<b>Research questions</b>	<b>22</b>
<b>3</b>	<b>Additions to Multilink</b>	<b>23</b>
3.1	Creating the lexicons . . . . .	24
3.2	Simulating different levels of L2 proficiency . . . . .	26
3.3	Determining the parameter value for the links between orthography and semantics . . . . .	27
<b>4</b>	<b>Simulating monolingual and bilingual word recognition</b>	<b>29</b>
4.1	Performance of Multilink versus the Interactive Activation model on 4 letter words . . . . .	30
4.2	Multilink's performance when orthography-semantic nodes are activated	31
4.3	Multilink's performance for all English words . . . . .	32
4.4	Multilink's performance for all Dutch words . . . . .	36
4.5	Performance of Multilink versus the Bilingual Interactive Activation model on 4 letter words . . . . .	40
4.6	Multilink's performance for all words in the bilingual lexicon . . . . .	41
4.7	Multilink's performance on cognates and controls . . . . .	44
4.8	Simulating word translation with Multilink . . . . .	45
<b>5</b>	<b>Conclusions and discussion</b>	<b>51</b>
	<b>References</b>	<b>56</b>

# 1 Introduction

It is very remarkable that we are able to acquire and use more than one language simultaneously, and that we can translate words and sentences from one language to another. In regions where multiple languages are spoken, this skill is of vital importance to economic, cultural, and scientific development. Translating words and sentences is so complex, because when a bilingual perceives a word (s)he “activates” word possibilities in both languages. Word translation implies that the bilingual is not only recognizing the word but also selecting the language to respond in and producing the proper response in that language. Even though word translation is a complex feat, all bilinguals, from young children to professional interpreters, are able to perform this task with ease. The field of psycholinguistics has gathered considerable knowledge about the component processes of translation in the past decades, but not much is yet known about the process as an integrated whole. In this thesis we will analyze the word translation process and describe its implementation in a computational model, as well as present simulations with this model, called Multilink, to clarify a number of current topics of research interest.

## 1.1 The complexity of word translation

One might think it would be easy to translate the English (L2) word HOOD (orthographic form) into the Dutch (L1) word /kap/ (phonological form) (La Heij et al., 1996). Just look up the meaning of ‘hood’ (translation recognition) and then produce the item /kap/ in the other language (translation production) (see Figure 1).

---

This thesis has lead to an article (Dijkstra et al., 2012) currently under review for *Bilingualism: Language and Cognition*.

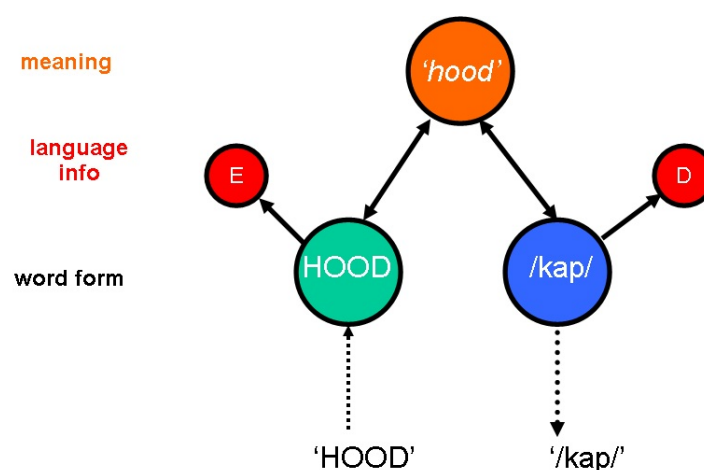
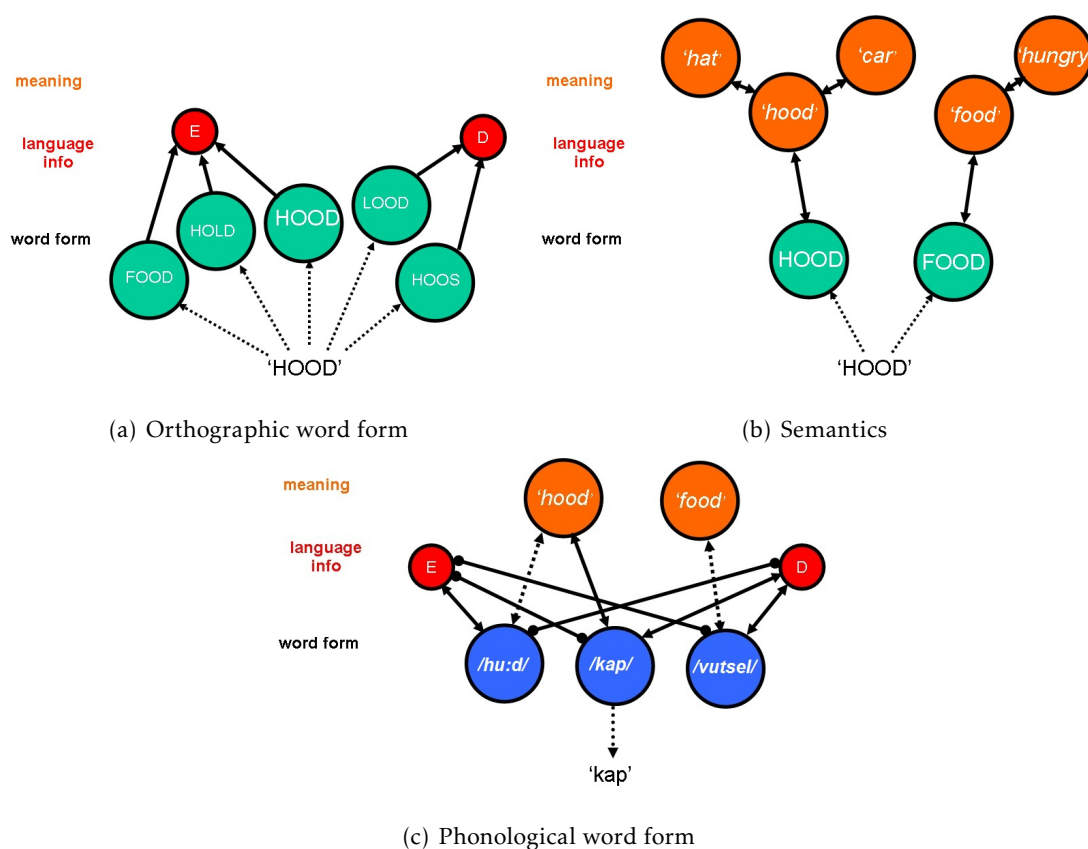


Figure 1: A visualisation of the word translation process.

But the translation process is actually much more complex. As can be inferred from the example, here a four-letter input must yield a three-phoneme output. Therefore the recognition and production processes must be able to handle the retrieval of words of different lengths. Also, depending on the context, at the meaning level the two words might not be fully equivalent conceptually. For instance, in particular contexts, */muts/* might be a better translation for HOOD than */kap/* (Tokowicz et al., 2002; Tokowicz and Kroll, 2007). Our research is focussed on the translation of words in isolation; the lack of context information therefore requires us to say that the “best” translation should come out first.

The process of word translation is further complicated by a process called “language non-selective lexical access”, i.e. the visual (or auditory) presentation of a word leads to a co-activation of many word candidates from different languages that are similar to the input (see Figure 2). For instance, in our running example the activated competitor set (or neighborhood or cohort) of HOOD will include the English words FOOD, HOLD, and HOOT, as well as the Dutch words LOOD, HOND, and HOOS. All these orthographic representations will then begin to activate their meaning representations (Balota, 1994; Grainger, 2008) in the second step of processing. Therefore, the presentation of HOOD will not only lead to the activation of the meaning *'hood'*, but the

meaning of words from the competitor set will in parallel be activated to some extent as well. Furthermore, semantically active representations will spread activation to other units. For instance, HOOD may spread activation to the meanings 'hat' or 'car', and FOOD to 'hungry'. Next, because the task is to produce a spoken output, semantically active representations will also activate phonological representations (in a language non-selective way), so the meaning 'hood' will activate its phonological representations /hu:d/ in English and /kap/ (for KAP) in Dutch. However, other active words will simultaneously activate *their* translations; for instance, the meaning of FOOD may also activate the Dutch phonological representation /vutsel/ for VOEDSEL.



**Figure 2:** A visualisation of language non-selective lexical access at the different levels of the word translation process.

As a result of the co-activation of several phonological representations of both the

input and the output language, a task-dependent decision process must make sure the correct translation is produced. For instance, it is likely that the most active phonological word candidate will be /hu:d/, because it was the input word. But this item may not be produced, because it belongs to the wrong language and it is not a translation from the input word. Another problem is that apart from the Dutch word /kap/, other Dutch words may be active, and therefore just selecting the most active Dutch word as a translation may not always succeed either. For instance, the English low-frequency input word JERK will co-activate its Dutch high-frequency neighbor WERK, which will directly activate its phonology /werk/ to considerable extent; if /werk/ is selected because it is Dutch and sufficiently active, the translation of JERK (i.e. /ruk/) will become available too late.

We also need to take into consideration cross-linguistic similarities and differences in vocabulary. For instance, we know that translation equivalents with form overlap (called ‘cognates’), like TOMATO (English) - TOMAAT (Dutch), lead to a facilitated translation process (Christoffels et al., 2006). The question is whether this effect arises on the recognition side or the production side of the translation process, or both.

Only a computational model will be able to take into account all these aspects of word translation at the same time. In order to implement such a model we have to be explicit about every assumption made with respect to each step of the translation process, no assumption remains implicit or hidden. But as a consequence this allows us to make both qualitative and quantitative predictions with respect to the time course of translation of words of different types by various groups of bilinguals and in all sorts of task situations.

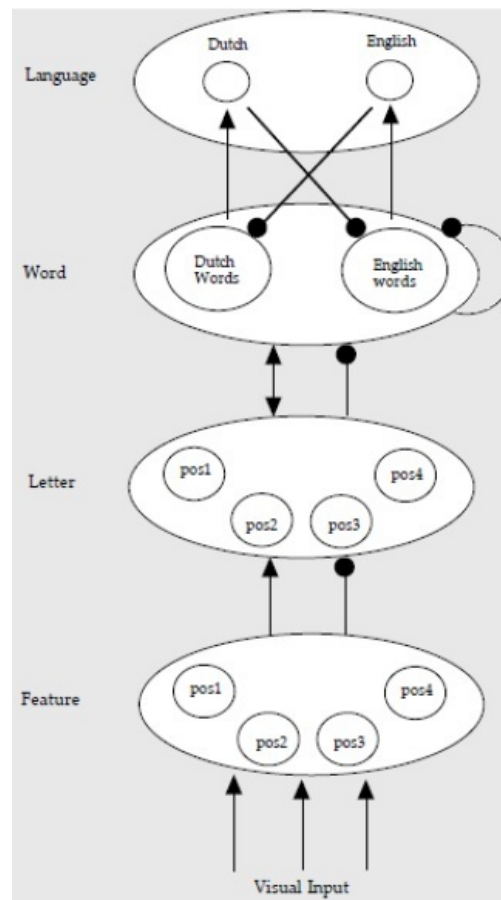
## 1.2 Models of bilingual word retrieval

As Figure 1 demonstrates, word translation consists of two components. Translation recognition takes place from an orthographic word form representation in one language to a meaning representation that is shared by all languages. Translation production proceeds from that meaning representation to a phonological word form rep-

resentation in another language. Several models have focussed on either the recognition component or the production component. One of the most influential computational models of (monolingual) visual word recognition is the Interactive Activation (IA) model (McClelland and Rumelhart, 1981). It is a localist-connectionist model, meaning that the symbolic units in the model represent objects in the world and that these representations are local (one unit per object) instead of distributed. It also implies that these units are connected to each other through excitatory or inhibitory connections (see McClelland and Rumelhart, 1981, for further details on the structure of the model). Although the IA model can account for many effects in the field of word recognition, it is not capable of dealing with multiple languages and it does not incorporate a word's semantic representation or a word's phonological representation.

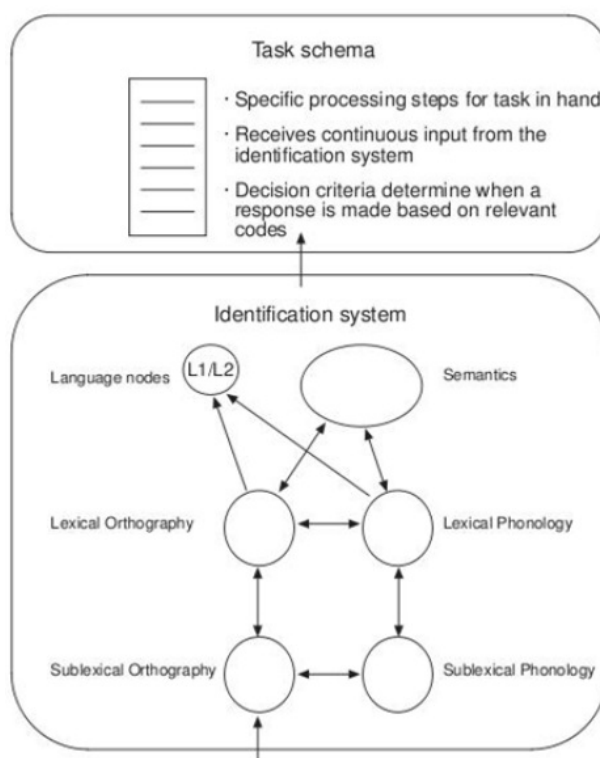
By extending the IA model to the bilingual domain, the Bilingual Interactive Activation (BIA) model (Figure 3) by Van Heuven et al. (1998) was able to account for the word recognition process of word translation. It implements bottom-up word activation in a language non-selective fashion (letters activate words from all languages). As a linguistic representation for language membership, language nodes were added to the original IA model structure. A language node receives activation from all orthographic word representations belonging to that language. This language node then inhibits the orthographic representations belonging to the other language. Although the BIA model can deal with an additional language, it is still limited by the lack of phonological and semantic representations.





**Figure 3:** The Bilingual Interactive Activation model (Van Heuven et al., 1998)

Later on, the BIA model was extended to the BIA+ model (Dijkstra and Van Heuven, 2002) that included lexically phonological and semantic representations of the words. In order to allow the model to execute different tasks and to distinguish task-related effects from linguistic effects a task/decision system was introduced (see Figure 4). The model proposes that the lexical activation levels within the word identification system itself are not affected by the task/decision system and, therefore, not by sources of non-linguistic information either.

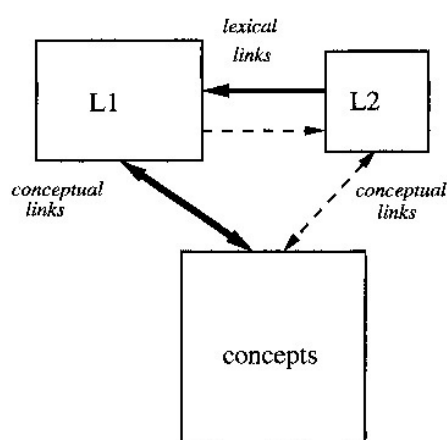


**Figure 4:** The BIA+ model (Dijkstra and Van Heuven, 2002)

A computational model that was able to account for the bilingual production component of word translation is the *WEAVER++* model (Roelofs, 1997). It is a computational model designed to explain how humans plan and attentionally control the production of spoken words. The *BIA/BIA+* and *WEAVER++* models can be considered as complementary to some extent. They are both symbolic in nature and make a distinction between a word retrieval system and a (not indicated) task/decision system, but where the *BIA/BIA+* models focussed on the word recognition process, the *WEAVER++* model focussed on the other component of word translation, the production component (see Roelofs (2002) for a comparison of *WEAVER++* and *BIA+*).

The Revised Hierarchical Model (RHM, see Figure 5) by Kroll and Stewart (1994) is a theoretical model that describes and examines word translation as a whole. However, it has never been actually implemented. It combines the word association model and the concept mediation model (Potter et al., 1984) into one model. Kroll and Stewart ar-

gued that most bilinguals know more words in the native than in the second language, as depicted by the larger box for L1 than L2. Also, lexical associations from L2 to L1 are assumed to be stronger than those from L1 to L2, because L2 to L1 is the common direction in which second language learners first acquire the translations of new L2 words. The links between words and concepts, however, are assumed to be stronger for L1 than for L2, which reflects the larger familiarity of bilinguals with L1 word meanings, as L1 is their native language. This is why, according to the model, a beginning L2 learner will often use an indirect road to translation, namely to link the L2 word form first to the L1 representation, which is then used to retrieve word meaning. Thus, this route asks for “word association”. At a later stage, when learners become more proficient in L2, the direct path from L1 or L2 orthographic representation to meaning, and then from meaning to L2 or L1 phonological representation, the “concept mediation” route, will become stronger and more important for word translation. According to the RHM, these considerations lead to longer response times for forward translation (L1 to L2) than for backward translation (L2 to L1).



**Figure 5:** Revised Hierarchical Model (Kroll and Stewart, 1994) of lexical and conceptual representation in bilingual memory

### 1.3 Empirical studies on word translation

The application of RHM is limited because it is a non-implemented theoretical model. This implies that the assumptions made in the model have never been properly (qualitatively and quantitatively) tested. However, several studies (both word recognition and word production studies) have attempted to reproduce the translation asymmetries predicted by the RHM, but these studies reported mixed results.

#### 1.3.1 Translation recognition

The translation recognition task is a forced-choice paradigm with two alternatives. A prime word in one language (e.g., BIKE) is presented for a short amount of time, and after a short delay it is followed by a target word (e.g., FIETS) in another language. Participants must then decide as quickly as possible (via a YES or NO response) whether the target is a correct translation of the prime. Most studies have focussed on the responses to incorrect translation pairs, such as pairs where the primes are word form or semantic distractors relative to the targets. Consider for example pairs like ANIMAL-NIER ('kidney') and ANIMAL-MUIS ('mouse') where the correct translation is ANIMAL-DIER. Responses to these target conditions are compared to those on unrelated item pairs like ANIMAL-DUIM ('thumb').

The translation recognition task has been used in several studies to study second language (L2) learning and translation. Talamas et al. (1999) studied how fluent and less fluent English-Spanish adult L2 learners handled form- or meaning-related translation pairs. They found that, when the results were summarized across translation directions, native language (L1) to second language (L2) and vice versa, the less fluent bilinguals appeared to be suffering more from word form interference than from semantic interference, while the exact opposite appeared to be true for the more fluent bilinguals. With the RHM in mind, Talamas et al. concluded that less fluent bilinguals rely on word form in word translation and more fluent bilinguals rely on word meaning.

By adding incorrect translation pairs with a target that was form-related to the prime

(like CARA - CARD) to the conditions of Talamas et al. (1999), and using a backward translation recognition task, Sunderman and Kroll (2006) obtained similar results for more or less fluent English-Spanish bilinguals. Less proficient bilinguals displayed both word form interference and semantic interference. More proficient bilinguals suffered from word form interference from the prime only and from semantic interference. When the translation pairs were derived from different grammatical classes the word form interference effect disappeared for all bilinguals.

Another study that suggests that non-proficient bilinguals rely more on form and more proficient bilinguals on meaning during the translation recognition process is the study by Ferré et al. (2006). They used several types of Spanish - Catalan translation pairs; correct translation pairs (like RUC - BURRO, 'donkey'), incorrect translation pairs with targets related to the correct translation in form (like RUG - RIEGO, 'watering'), were close in meaning (like RUC - OSO, 'bear'), or were very close in meaning (like RUC - CABALLO, 'horse'). When the results were compared to unrelated controls, both early proficient bilinguals (bilinguals who were proficient in both languages at an early age) and late proficient bilinguals (bilinguals who were proficient in both languages at a late age) showed more interference from meaning than from word form distractors, but only from distractors that were closely related in meaning to the correct translation, while late non-proficient bilinguals showed more interference from form than from meaning distractors.

There are also studies which conclude that both less proficient bilinguals and more proficient bilinguals experience word form and semantic interference effects when translating words from L2 to L1, for instance, the study by Altarriba and Mathis (1997) where English monolingual university students learned Spanish-English translations. Next, these L2 learners and a group of fluent English-Spanish bilinguals performed a translation recognition task, in which both groups showed word form and semantic interference effects.

A comparison was made between the brain activity and response behaviour of Dutch primary school children, who had just begun to learn English words, to that of pro-

ficient adult Dutch-English bilinguals in a backward translation recognition task, by Brenders et al. (under revision). As in other studies, the incorrect translation pairs consisted of semantic distractors, word form distractors and incorrect control pairs. Reaction times, errors and ERPs (Event-Related Potentials) all indicated the presence of form and meaning activation in all participating groups.

Another interesting result was obtained by Comesaña et al. (2009). Comesaña et al. used two different teaching methods to teach children L2 vocabulary; L2-picture association learning or L2-L1 association learning. For L2-picture association learning, even after just one vocabulary learning session, a significant semantic interference effect was obtained in a translation recognition task. This finding suggests that L1 word form representation does not always mediate vocabulary acquisition but that children quickly create a conceptual representation of L2 words. Also, the semantic interference effect after learning through L2-L1 word association was non-significant. It would therefore appear that the obtained effects depend on the vocabulary teaching method that was used. Possibly, the method affects the type of links that are created in the initial stage of L2 learning. It would appear that in some contexts, the translation process proceeds (also) via the semantics, not (only) via word form links.

So, although not all effects are found in all studies, both word form and conceptual interference have regularly been observed in beginning and proficient L2 users, using the translation recognition task. While factors like teaching method and learning environment may modulate the result patterns, it appears to be a safe conclusion that already in an early stage of L2 learning, conceptual information is used in translation recognition. This is already an indication that the word association route via word form links suggested in the RHM may be problematic.

### 1.3.2 Translation production

Not only translation recognition studies reported mixed results. We also note mixed results in translation production studies. In a translation production task a participant is shown a word, and is then asked to produce the correct translation of that word as

fast as possible. The translation production task is often compared to production tasks such as picture naming, word naming, or Stroop (Dyer, 1973), because it has often been used to study how L2 word forms are mapped onto meaning.

Based on the RHM, it has been proposed that there is an asymmetry in translating from L1 to L2 and vice versa. In their study Kroll and Stewart (1994) found that response times for translation from L2 to L1 (backward translation) were shorter than from L1 to L2 (forward translation). Other studies have reported similar results (Francis and Gallard, 2005; Miller and Kroll, 2002). However, equally long or even faster response times for L1 to L2 translations were found for a number of other translation studies (Christoffels et al., 2006; De Groot and Poot, 1997; Duijck and Brysbaert, 2004; La Heij et al., 1996). We will now discuss a number of these studies.

La Heij et al. (1990) studied how L1 (Dutch) distractor words affect L2 (English) to L1 translation. Subjects were shown an English word to be translated. This word was quickly replaced by a Dutch distractor word that was semantically related or unrelated to the correct response. The authors found that translation took longer in the related than in the unrelated condition. In a later study (La Heij et al., 1996), pictures were presented instead of distractor words in a similar fashion. In four experiments, the relatedness effect proved to be just as large in both translation directions or larger when translating from L2 to L1 (further see Bloem and La Heij, 2003; Bloem et al., 2004; Miller and Kroll, 2002; Sholl et al., 1995).

Several studies show that semantic variables also seem to play a role in the different response times between forward ( $L1 \rightarrow L2$ ) and backward ( $L2 \rightarrow L1$ ) translation. De Groot et al. (1994) found that meaning variables affected both backward and forward translation, although they played a smaller role in backward than in forward translation. By examining how Dutch native speakers at different proficiency levels in L2 English translated words, De Groot and Poot (1997) found that proficient bilinguals produced equally long response times in both translation directions, while low proficient bilinguals even produced shorter response times in L1 to L2 translation. They observed that these results were affected by the semantic variable concreteness (whether

the translation pairs are semantically concrete or abstract). Another semantic variable of influence appeared to be number magnitude in both translation directions. Duijck and Brysbaert (2004) showed that translating number words for large quantities took longer than translating number words for small quantities.

There is still some debate whether the results of these studies can be explained by assuming that fluent bilinguals are not only able to map L1 word forms but also L2 word forms straight onto meaning, or whether they could be due to faster access to a well-known language than to a less well-known language and the difference between recognition and recall. An implemented word translation model can clarify this issue. It can test if the asymmetric effect predicted by the RHM does indeed occur and to what extent differences between recognition and production are responsible for it.

In regard to the assumptions made in the RHM, a few translation production studies are of particular interest to us. De Groot et al. (1994) and De Groot (1992) reported faster backward translation than forward translation, or just as fast, for low proficient bilinguals. Kroll et al. (2002) also reported faster backward translation, and they found this asymmetry effect to be larger in low proficient bilinguals. These results are in line with the asymmetry predictions made in the RHM.

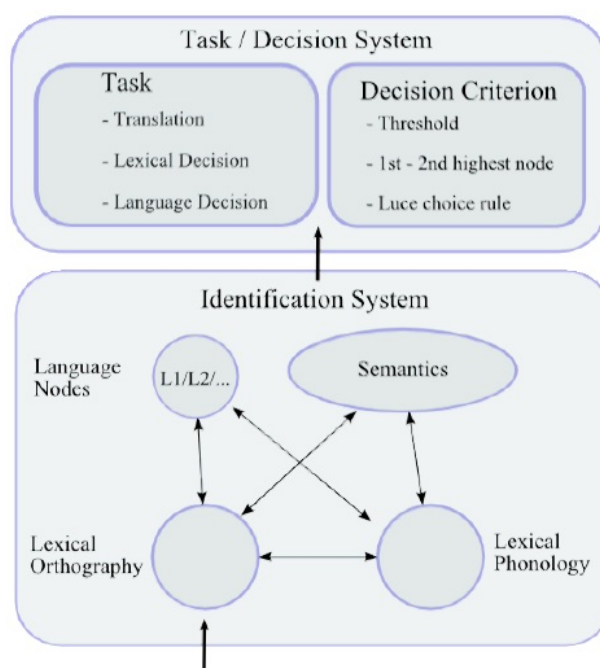
However, Christoffels et al. (2006) compared the word translation performance of trained interpreters to that of bilingual university students and highly proficient English teachers. The authors also reported a larger asymmetry effect for low proficient bilinguals than for high proficient bilinguals, but they differed in the direction of this asymmetry. Low proficient bilinguals translated significantly faster in forward translation. A summary of the reaction time results in these studies is provided in Table 13 in the word translation section of our paper.

In our section about the word translation simulations with Multilink we will compare Multilink's performance on the asymmetry predictions made in the RHM with the results of these studies.



## 1.4 Multilink

In the previous sections it has become clear that there are still some issues regarding the word translation process that need to be clarified. An implemented computational model of word translation has the potential to clarify some of these issues. This is why Dijkstra and Rekké (2010) introduced Multilink. Multilink has an underlying architecture that combines aspects of the IA model, BIA/BIA+ models, WEAVER++ and the RHM (see Figure 6 for Multilink's architecture). The representational architecture for word translation, the identification system, consists of an integrated multi-language lexicosemantic system as proposed by BIA/BIA+ and WEAVER++, containing orthographic, semantic and phonological representations for multiple languages. Just like these models, Multilink also has a task/decision system implemented in order to take care of task- and context-sensitive decision processes.



**Figure 6:** Multilink's architecture (Dijkstra and Rekké, 2010)

Multilink describes the word translation process as a whole. It is the first computational model of word translation that models word translation in the way humans

perform the task. Dijkstra and Rekké have implemented the model in such a way that it is able to overcome the complicating factors of word translation mentioned earlier in this thesis. In the activation of stored lexical representations by input words a modified Levenshtein distance is used (see Section 1.4.2) that enables the model to handle words of different lengths. It is also able to handle words with cross-linguistic similarities and differences, like cognates (words with form and meaning overlap across languages) and false friends (words with only form overlap across languages). Further, to overcome the problem of language non-selective lexical access Dijkstra and Rekké have implemented a task/decision system. This task/decision system allows Multilink to model performance in different tasks, such as lexical decision, language decision, and word translation recognition/production. Multilink supports two types of lexical decision. The first is standard lexical decision, in which the model decides whether a word belongs to a particular language or not, here the output is a YES or NO response and the timestep on which the decision was made. The second is generalized lexical decision, in which the model decides whether the input word is a word at all, also with a YES or NO response and the timestep on which the decision was made. In the language decision task the model has to decide to which language an input word belongs. Here the output is the name of the winning language node and the timestep upon which the decision was reached. In the word recognition task the output is the first word that crossed the recognition threshold, together with the timestep upon which this threshold was crossed. Crossing the threshold in this case means that the representational unit in the network representing this word has received a total activation of 0.7 (70% of its maximal value) or more, the same threshold value as originally used in the IA and BIA/BIA+ models.

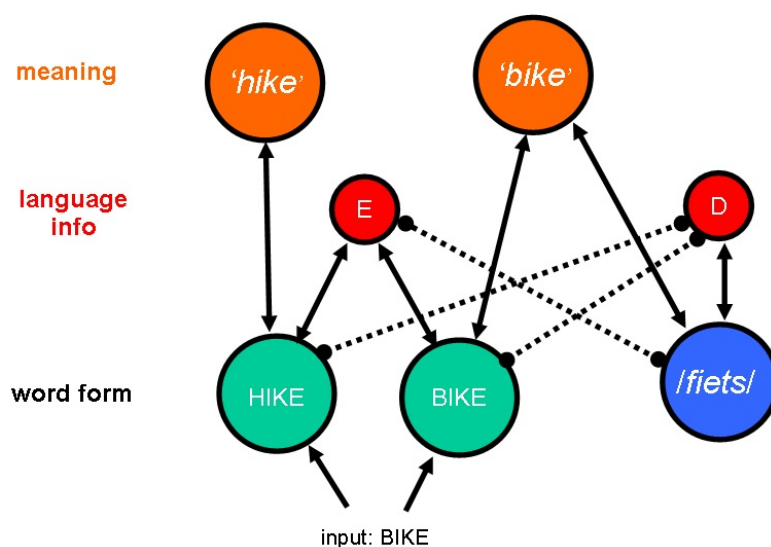


Figure 7: Multilink's identification system

The word identification system (shown in Figure 7) is a localist connectionist network containing several layers of symbolic representations, like the IA and BIA models. This figure demonstrates that Multilink is a localist-connectionist model because the representations are local (i.e. one unit per “object”) as opposed to distributed, and the units are connected to each other through excitatory (normal arrows) or inhibitory (interrupted lines) connections. Figure 7 shows that Multilink does not only include the orthographic activation based on the input string, but also activation resulting from feedback from linked semantic representations. On every time step of model simulation, the orthographic representations activate their linked semantic representations, which in turn resonate with the orthographic representations on a one-by-one basis. Multilink differs from the AI/BIA models in this aspect, because those do not include semantic representations and can therefore only capture orthographic activation based purely on the input.

What is furthermore important to note about Multilink is that it has been implemented in such a way that it includes both the word association route and the concept mediation route as suggested by the RHM. In order to refute the necessity of the assumptions made in the RHM, in the version of Multilink we used in this thesis, we

have set the weights of the word form links between the L1 and L2 word form representations to zero. This means that there can be no spreading of activation via word form links, so there will be no involvement of the word association route in the results obtained by us with this version of Multilink.

#### 1.4.1 Activation functions and parameter settings

As shown in Figure 7, the word identification system of Multilink is very similar to the IA and BIA models discussed earlier in Section 1.2. It is also a localist-connectionist network containing several layers of symbolic representations. Just as other localist-connectionist models, Multilink assumes that word representations have a resting level activation that depends on its frequency of usage. This means that the network gives high frequency words a headstart in the recognition process, just as people recognise high frequency words faster. To determine the resting level activation of each word in a particular language, all words were ranked from high to low frequency and then scaling factors were applied, using the following resting level activation formula:

$$MIN\_REST + RANK * \frac{MAX\_REST - MIN\_REST}{MAX\_RANK} \quad (1)$$

Here  $MIN\_REST$  is the parameter that determines the minimal resting level activation (set at -0.05).  $RANK$  is the ranking of the word in the frequency ordered concept list (from high frequency to low frequency) for that particular language.  $MAX\_REST$  is the parameter that determines the maximal resting level activation (set at 0.00). And  $MAX\_RANK$  is the highest rank that occurs in the frequency ordered concept list. Words of the same language and of the same frequency have the same  $RANK$  and Equation 1 will therefore give them the same resting level activation.

For the spreading of activation throughout the network, Multilink uses the same activation functions as the IA model and the BIA/BIA+ models. Equation 2 shows how the net input ( $n_i(t)$ ) for a node is computed at a given timestep  $t$ ,

$$n_i(t) = \sum_j \alpha_{ij} e_j(t) - \sum_k \gamma_{ik} i_k(t) \quad (2)$$

where  $\alpha_{ij}$  is the weight of an incoming excitatory connection between node  $i$  and neighboring excitatory node  $j$  ( $\alpha_{ij} = \alpha_{ji}$ ),  $e_j(t)$  is the activation of node  $j$  at timestep  $t$  (where  $e$  stands for excitatory),  $\gamma_{ik}$  is the weight of an incoming inhibitory connection between node  $i$  and neighboring inhibitory node  $k$  ( $\gamma_{ik} = \gamma_{ki}$ ) and  $i_k(t)$  is the activation of node  $k$  at timestep  $t$  (where  $i$  stands for inhibitory). Depending on whether the net input is positive or negative, it can have a different contribution to the effect  $\epsilon_i(t)$  on a node, as shown by Equation 3 below, where  $M$  is the maximum activation of a node (set at 1.0),  $m$  is the minimal activation of a node (set at -0.2) and  $\alpha_i(t)$  is activation of a node at timestep  $t$ .

$$\epsilon_i(t) = \begin{cases} n_i(t)(M - \alpha_i(t)) & \text{if } n_i(t) > 0 \\ n_i(t)(\alpha_i(t) - m) & \text{if } n_i(t) \leq 0 \end{cases} \quad (3)$$

The activation is then calculated with the activation function shown in Equation 4, in which  $\Theta_i$  is the decay rate of the node (set at 0.07, just as in the IA/BIA/BIA+ models) and  $r_i$  is the resting level activation of the node (determined by Equation 1).

$$\alpha_i(t + \Delta t) = \alpha_i(t) - \Theta_i(\alpha_i(t) - r_i) + \epsilon_i(t) \quad (4)$$

To put it more plainly, by adding the current net input from all other connected nodes, including the presented word, to its activation at the previous time step, the activation of the lexical representation is updated at each time step. Depending on the degree of orthographic overlap between the stored representation and the presented word, each stored representation receives activation from the input word to a proportion of its similarity strength. This means that a number of word candidates are activated to varying extent by the input, depending on the orthographic similarity with the input and on their frequency. For a more detailed description of Multilink we refer to Dijkstra and Rekké (2010).

#### 1.4.2 The modified Levenshtein distance

Where the IA and BIA models were only able to handle words of a fixed length, Multilink is able to handle input words of different lengths by using a modified Levenshtein

distance (Schepens et al., 2011). In this thesis we use 3-8 letter words. The formula to determine the modified Levenshtein distance is as follows:

$$score = 1 - (distance/length) \quad (5)$$

where  $length = \max(\text{length of source expression}, \text{length of destination expression})$  and  $distance = \min(\text{number of insertions, deletions and substitutions})$ .

By computing a modified Levenshtein distance between the involved letter strings the orthographic similarity between the input word and the stored lexical representations is determined. The Levenshtein distance involves the computation of the minimal number of deletions, substitutions, and insertions needed to edit one expression into the other. In the standard Levenshtein distance the length of the expressions is not taken into account, so two word pairs with an equal number of mismatching letters are considered to have an equal distance, while in the modified version two pairs are considered to have an equal distance if the ratio of overlap is the same. For example, in the modified Levenshtein distance measure the pair “round” and “bound” gets a score of 4/5 while the pair “forces” and “forced” gets a score of 5/6 even though they would have the same standard Levenshtein distance.

## 2 Research questions

Dijkstra and Rekké (2010) were successful in implementing a computational model for word translation, but they have not yet been able to extensively test this model. The lack of a sufficient amount of lexical data meant that they could only demonstrate the presence of empirical effects found in the domain of psycholinguistics by limiting themselves to examples. The main research question in our research is therefore:

- *Can we validate Multilink based on empirical data?*

In the monolingual domain we will first perform a model-to-model comparison and a model-to-data comparison in order to validate Multilink. In these simulations we will try to answer the following research questions:

- *How does Multilink perform compared to a computational model for monolingual word recognition, the Interactive Activation (IA) model?*
- *Can Multilink reproduce the empirical effects found in the psycholinguistic domain?*

If we can validate our model with these simulations we will extend to the bilingual domain, in which we will perform another model-to-model comparison and some more model-to-data comparisons to answer some more research questions:

- *How well does Multilink perform compared to the Bilingual Interactive Activation (BIA) model and can Multilink also reproduce the empirical effects in bilingual simulations?*
- *Can Multilink reproduce the translation latencies for cognates and non-cognates as observed in empirical studies?*
- *How will Multilink perform on the general asymmetry predictions for forward and backward word translation made by the RHM?*

In the following sections we will try to answer these questions. We will first describe some additions we had to make to Multilink in order to answer these questions. Then we will discuss our monolingual and bilingual simulations that served to validate Multilink, we will discuss how Multilink performs on words with cross-linguistic similarities and differences, and we will test Multilink's performance on the asymmetry predictions made by the RHM. We will end this thesis with a discussion of the results and the implications that these results have.

### 3 Additions to Multilink

In order to answer our research questions we first had to create some additions to Multilink. Dijkstra and Rekké (2010) did not have an extensive amount of lexical data available to test their model with. Our first addition was to create the lexicons (two monolingual lexicons, one Dutch one English, and a bilingual Dutch-English lexicon)

which we will use for our simulations. To properly test Multilink's performance on the asymmetry predictions made by the RHM, Multilink has to be able to simulate different levels of L2 proficiency. Finally, to test the influence of adding semantics, we had to find a parameter value for the weight of the links between orthographic word form representations and semantic representations that resulted in Multilink producing high correlation values with empirical data as well as perform well on a semantic priming task.

### 3.1 Creating the lexicons

We had to create Dutch and English monolingual lexicons as well as a bilingual Dutch-English lexicon. In order to answer our research questions we need to know for each word, its frequency of usage in a language, the number of letters of that word, a target word which is semantically related to the input word, an indication of how strongly the input word is associated to the target word, the average time it takes people to recognize the word, and a percentage indicating how often the word is correctly recognized by people. Table 1 summarizes the types of lexical data we want to have in our lexicons.

Datatype	Description
word	the word which will serve as input
wf	word frequency, in frequency per million
#l	the length of the word
target	target word, a word that is semantically related to the input word
wass	word association strength, a value between 0 and 1 which describes how strongly a word is associated to the target word
RT	reaction time, the average time it takes people to recognize the word
%acc	recognition accuracy, percentage of time the word is correctly recognized by people

**Table 1:** Language information that needs to be present in a lexicon.

Below we describe how we created those lexicons. First the English monolingual lexicon was created, by combining the contents of three databases: (1) the CELEX database



by Baayen et al. (1993), which contains frequency data for English words in occurrences per million; (2) the Free Association Database created by Nelson et al. (1998), which contains the semantic relations between concepts; and (3) the database of the English Lexicon Project, which contains the average reaction times and accuracy values for a collection of English words (Balota et al., 2007). By cross-referencing these databases to each other we extracted all the 3-8 letter English words together with the lexical data, thereby creating an English monolingual lexicon consisting of 4017 words.

Each word in the English lexicon was then paired with its Dutch translations according to a word translation database (Linguistic Systems B.V., 2008). Translation often results in words of different lengths. Since we restrict our experiments to words of 3-8 letters in length, we only kept the first Dutch translation that meets this criterium. This means that certain Dutch translations are not eligible to act as an English-Dutch translation pair, for instance, a pair like “advance” - “voorschot” was removed. The resulting list of Dutch words lacks response time and association data. Response time data from the Dutch Lexicon Project by Keuleers et al. (2010) were therefore added to the Dutch items. This resulted in a Dutch monolingual lexicon consisting of 2180 Dutch words with response time data. For 595 of these words Dutch association data was found in the corpus by De Deijne and Storms (2008). This means that we only have Dutch word association data available for a fraction of our Dutch words, whereas we have English association data available for all 4017 English words. We therefore created two Dutch monolingual lexicons, one containing 2180 Dutch words without word association data, and one containing 595 Dutch words with association data. Given that we only have word association data available for a fraction of our Dutch words, we decided not to use the word association data in our simulations. This means that we only used the Dutch lexicon containing 2180 words without word association data. The smaller Dutch lexicon with association data can possibly be used for future research.

The bilingual lexicon was then created by combining the Dutch and the English lexicons. This resulted in a lexicon containing 2180 Dutch-English translation pairs (without association data). Multilink acts under the assumption that Dutch and English

words are stored in one abstract memory system, meaning that they should be represented in an integrated lexicon. The created lexicons contain several kinds of words, identical and non-identical cognates, false friends and control items. In our simulations, the parameter regulating lexical competition was set to 0, therefore, depending on the task at hand, only response competition can affect processing times and lateral inhibition can not.

### 3.2 Simulating different levels of L2 proficiency

As mentioned before in Section 1.4.1, Multilink assumes that word representations have a resting level activation that depends on its frequency of usage. This means that the network gives high frequency words a headstart in the recognition process, just as people recognise high frequency words faster. To determine the resting level activation of each word in a particular language, all words were ranked in a frequency ordered concept list and then scaling factors were applied. To properly test Multilink's performance on the asymmetry predictions made by the RHM, we must first adapt Multilink so that it is able to simulate different levels of L2 proficiency.

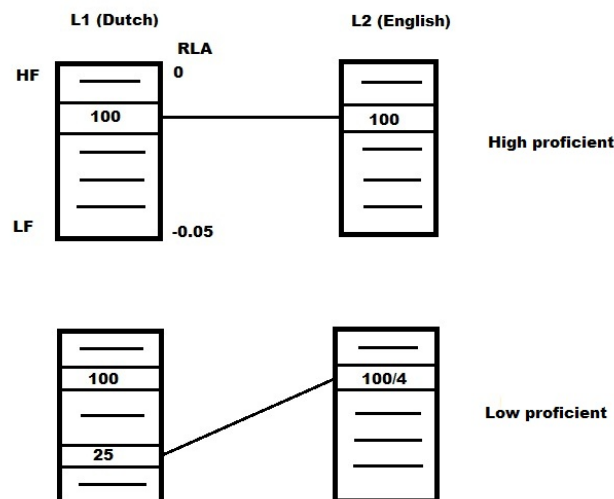
We first determined the L1 (Dutch) resting level activations. All the Dutch words were ranked based on their frequency, thereby creating a frequency ordered concept list. Based on a word's rank in this frequency ordered concept list, and on Equation 1 in Section 1.4.1, each word was then given a resting level activation between the values of 0.00 (maximal resting level activation value) and -0.05 (minimal resting level activation).

Figure 8 depicts how we then simulated different levels of L2 proficiency. We first determined the resting level activation values for high proficient bilinguals, by creating a frequency ordered concept list of all the L2 (English) words, and then giving an English word of a certain frequency the same resting level activation value as was given to a Dutch word of the same frequency.

For low proficient bilinguals, the English frequency obviously must be lower than that for native English speakers. Currently it is not known to what extent and in what

way. Therefore, we decided to approximate the English frequency in unbalanced bilinguals by dividing the native frequency by 4. This means that for low proficient bilinguals, we gave an English word with a frequency of 100 occurrences per million the same resting level activation as a Dutch word with a frequency of 25 occurrences per million.

By determining the resting level activation values this way, the Dutch words in a bilingual lexicon are not affected by the number and subjective frequencies of English words in the bilingual lexicon. This gives us the possibility to lower the frequency of English words and adapt the associated resting level activation values, without influencing the resting level activation values of Dutch words.



**Figure 8:** A graphic interpretation of the way different L2 proficiency levels are simulated.

### 3.3 Determining the parameter value for the links between orthography and semantics

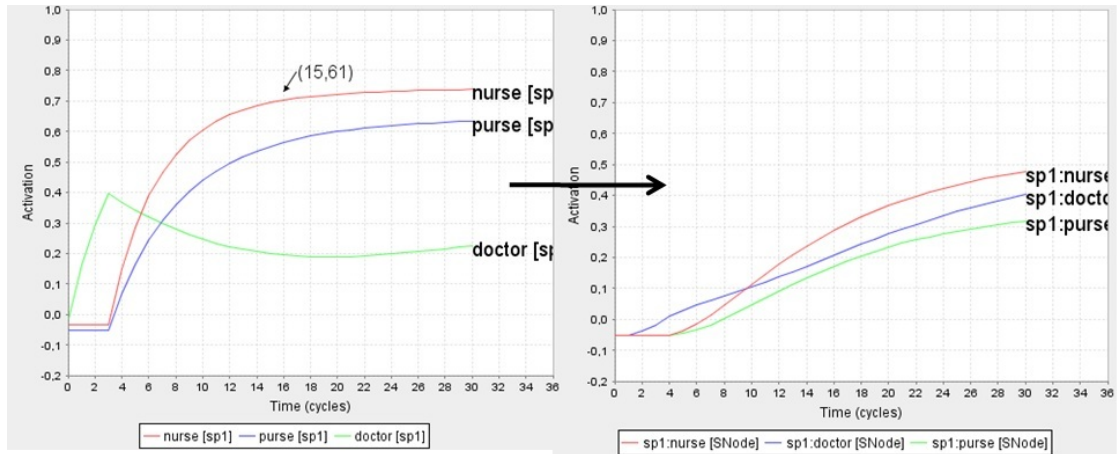
Multilink is implemented in such a way that the orthographic word form representations do not only receive activation based on the similarity with the input word, but also receive activation due to feedback between orthographic word from representations and semantic representations. Since Multilink is the first computational model to

model word translation the way humans do it, there is no reference as to what would be a good value for the weights regulating this feedback. We therefore had to determine this value ourselves.

We set out to find a value for the parameters regulating the flow of activation from the orthographic word form representations to the semantic representations (the OS parameter), and from the semantic representations to the orthographic word form representations (the SO parameter). We wanted to find a value for these parameters that would allow Multilink to produce high correlations with reaction times in a word recognition task, as well as resulting in a good model performance in a semantic priming task.

To simulate human performance in a semantic priming task, a prime is first presented to the model at timestep  $t = 0$ , e.g. the word DOCTOR. Then a few timesteps later, say  $t = 3$  a target word and a control word are presented to the model. The target word is semantically related to the prime, e.g. NURSE, while the control word has no relation to the prime, but is orthographically related to the target word, e.g. PURSE.

A good model performance in a semantic priming task would then mean that at the orthographic level NURSE received the highest activation, followed by PURSE because it has a high orthographic similarity to NURSE, and DOCTOR should receive the lowest amount of activation because it is not orthographically related to NURSE or PURSE. However, at the semantic level, NURSE should receive the highest activation, because it is the target word, followed by DOCTOR, because it is semantically related to the target word, followed by PURSE, which has no semantic relation to the other two words. Figure 9 shows that we are able to demonstrate this effect with Multilink.



**Figure 9:** A demonstration of the semantic priming task performed with Multilink. The graph on the left represents the activation at the orthographic level, while the graph on the right represents the activation at the semantic level.

We used a trial-and-error approach to determine Multilink’s correlation with reaction times in a word recognition task and Multilink’s performance in a semantic priming task, for several values for the OS and SO parameters. In this way we determined that giving both parameters a value of 0.03 gave the best performance in both tasks.

## 4 Simulating monolingual and bilingual word recognition

With the additions described in Section 3 we can validate Multilink. In the following sections we will discuss a number of monolingual and bilingual simulations. With the first simulations we will validate Multilink in the monolingual domain. Then we will repeat these simulations in order to validate Multilink in the bilingual domain. Next, we will discuss the simulations meant to investigate Multilink’s performance on words with cross-linguistic similarities and differences. Finally, we will discuss word translation simulations with Multilink to test Multilink’s performance on the asymmetry predictions made by the RHM.

#### 4.1 Performance of Multilink versus the Interactive Activation model on 4 letter words

For this simulation we performed a model-to-model comparison between Multilink and the Interactive Activation (IA) model (McClelland and Rumelhart, 1981), which is a computational model for word monolingual word recognition. There are two limitations to the IA model that must be mentioned here. First, the IA model can only deal with words of a fixed length, e.g. only 4 letter words or only 5 letter words. Most studies on monolingual word recognition with the IA model use 4 letter words, therefore we fed all the 4 letter words in the English and Dutch lexicons one by one to Multilink and to the IA model. Second, the IA model is a model of word recognition and does therefore not include any semantic representations. This means that for a true comparison between the two models, the activation resulting from feedback between the orthographic word form representations and the semantic representations in Multilink should not be taken into account. In other words, the OS/SO parameters should be set to 0.

Given the types of lexical data we have available in our lexicons, it is interesting to compare the resulting correlations of the two models with the reaction times, word frequency and the logarithm of the frequency. In total, 850 English and 470 Dutch 4 letter words were used for the simulations. The resulting correlations of the number of cycles of Multilink and the IA model with reaction times (RTs), word frequencies (Freq) and the logarithm of the word frequencies (LogFreq) are given in Table 2.

	English N=850			Dutch N=470		
	RTs	Freq	LogFreq	RTs	Freq	LogFreq
Multilink	.423	-.462	-.978	.584	-.368	-.976
IA	.274	-.308	-.575	.203	-.177	-.429

**Table 2:** Correlations of the number of cycles in Multilink and the IA model with reaction times (RTs), word frequency (Freq), and the logarithm of the word frequencies (LogFreq) for English and Dutch four letter words. Orthographic-semantic links set at 0.00.

Table 2 shows that Multilink consistently performed better than the monolingual IA model did. The high correlations of Multilink's cycle times with the logarithm of the frequency are the result of the fact that Multilink's resting level activations are directly determined by frequency ranking. However, the correlation of IA and word frequency is relatively low and may also be due to the indirect way the IA model converts rank order into resting level activation. Table 2 also shows that Multilink produces a very good correlational fit to the empirical reaction time data. The correlation of the number of cycles with the reaction times (RTs) is twice as high for Multilink than it is for the IA model. Finally, the reported data indicate that the correlations are generalizable across languages and provide support for Multilink.

#### 4.2 Multilink's performance when orthography-semantic nodes are activated

Multilink is intended as a model for the word translation process as a whole, not just the word recognition component. To simulate the word translation process as a whole, it needs to include semantic representations. To see what would happen when feed forward and feed back between orthography and semantics is introduced, we repeated the previous series of simulations, this time setting the OS/SO parameters to 0.03. The results are shown in Table 3.

	OS/SO = .03 English N=850			Dutch N=470		
	RTs	Freq	LogFreq	RTs	Freq	LogFreq
Multilink	.365	-.543	-.940	.493	-.447	-.953
IA	.274	-.308	-.575	.203	-.177	-.429

**Table 3:** Correlations of the number of cycles in Multilink and the IA model with reaction times (RTs), word frequency (Freq), and the logarithm of the word frequencies (LogFreq) for English and Dutch four letter words. Orthographic-semantic links set at 0.03.

As can be concluded by comparing Multilink's correlations in Table 3 with those in Table 2, incorporating semantics in Multilink does not lead to any serious deterioration

of model performance. Multilink's correlation values in Table 3 are slightly lower than in Table 2. Essentially, with the OS/SO parameters set to 0.03, we included the semantic representations in the word recognition task, thereby giving more information to the model. More information also leads to more noise. However, despite these correlation values being slightly lower, Multilink still outperforms the IA model.

### 4.3 Multilink's performance for all English words

Multilink was designed to be able to handle words of varying lengths as input. Therefore it is very interesting to see how well Multilink performs with words of different lengths. In this series of simulations we compared Multilink's performance with both the OS/SO links turned "off" and turned "on". The resulting correlations with reaction times for 4017 English words of different lengths are shown in Table 4.

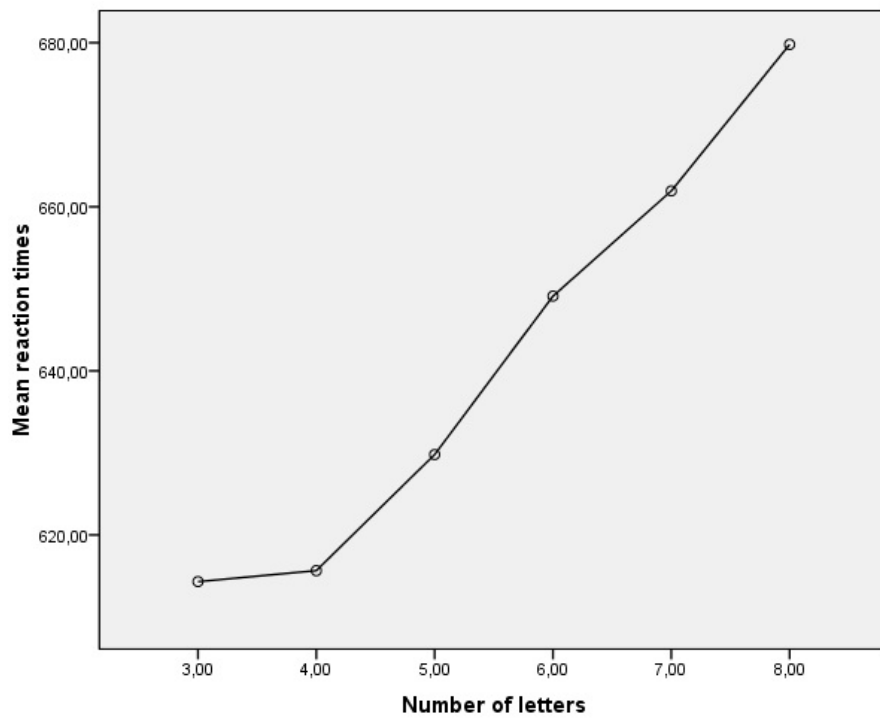
	all	3	4	5	6	7	8
OS/SO=0	.423	.514	.423	.466	.458	.447	.412
OS/SO=0.03	.360	.454	.365	.413	.411	.406	.383

**Table 4:** Correlations of the number of cycles in Multilink with reaction times to English words of different lengths. Orthographic-semantic links set to 0 or 0.03.

In both sets of simulations, the correlations computed on the basis of a large number of items (see Table 5) are high.

An empirically established effect in visual word recognition is the word length effect, which claims that longer words take longer to recognize. Figure 10 displays a graph plotting the average reaction times in our English lexicon against the number of letters. Here we see that the word length effect indeed occurs in our lexical data. The graph clearly shows that with increasing word length, the mean reaction times increase as well. Because we have words at our disposal ranging from 3 to 8 letters in length, we also examined the word length effect in Multilink. Table 5 shows the result for OS/SO=0 and Figure 11 shows the result pattern in graph form.

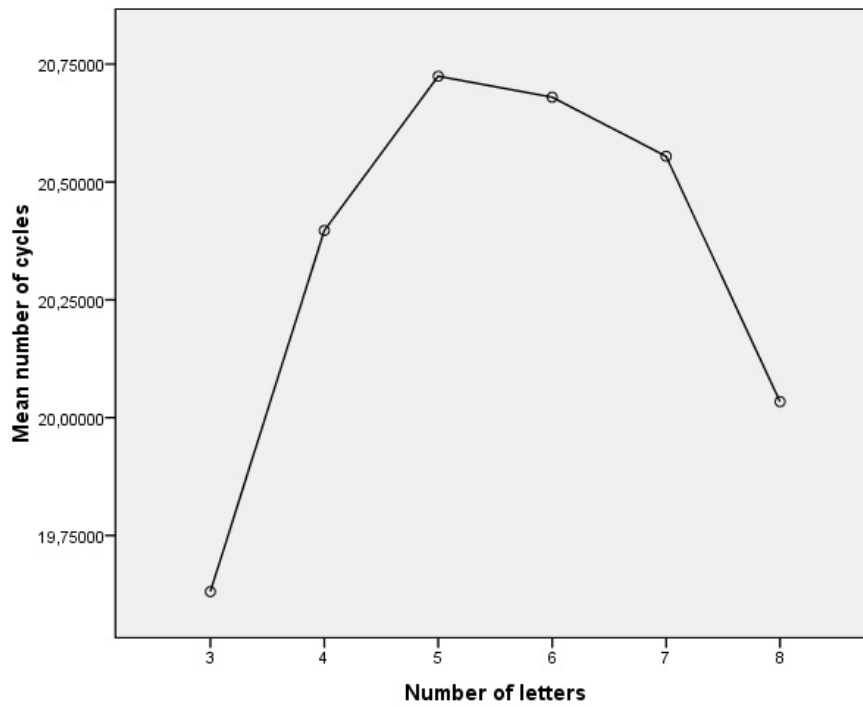




**Figure 10:** Word length effect for all our English words, mean reaction times plotted against number of letters

Number of letters	Mean	Standard deviation	N
3	19.63	3.58	283
4	20.40	3.42	850
5	20.72	3.37	935
6	20.68	3.37	829
7	20.55	3.44	678
8	20.03	3.48	442
Total	20.46	3.43	4017

**Table 5:** Mean number of cycles in Multilink per word length for OS/SO=0 for English words.

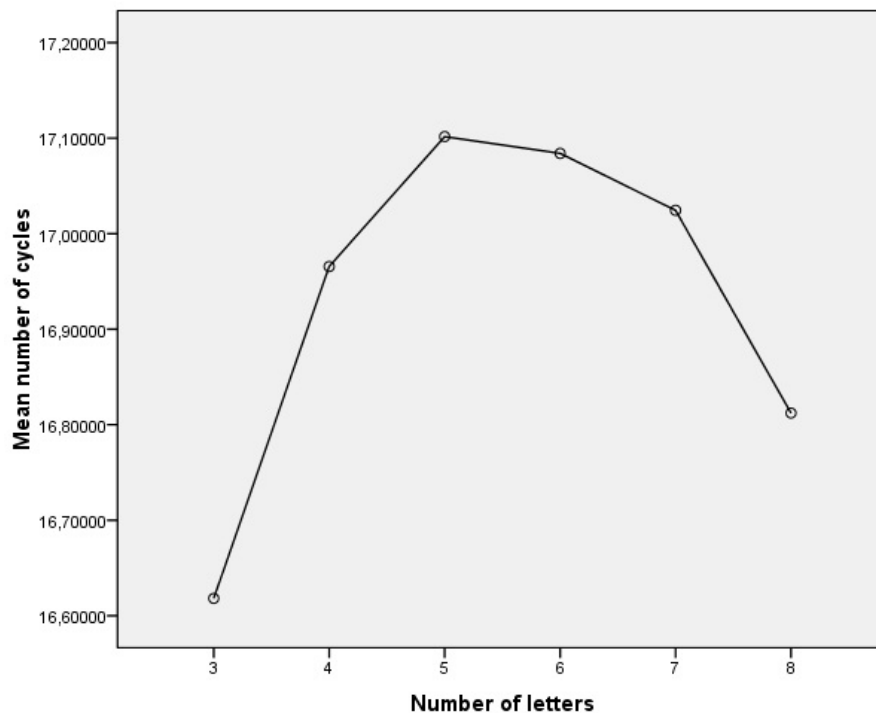


**Figure 11:** Word length effect English OS/SO=0, mean number of cycles plotted against number of letters

And for OS/SO=0.03 the results are shown in Table 6 and the pattern of these results can be seen in Figure 12.

Number of letters	Mean	Standard deviation	N
3	16.62	1.62	283
4	16.97	1.47	850
5	17.10	1.41	935
6	17.08	1.41	829
7	17.02	1.44	678
8	16.81	1.50	442
Total	16.99	1.46	4017

**Table 6:** Mean number of cycles in Multilink per word length for OS/SO=0.03 for English words.



**Figure 12:** Word length effect English OS/SO=0.03, mean number of cycles plotted against number of letters

We expected to see an increasing line in the graphs, since the word length effect predicts a higher number of cycles for longer words. Figure 11 and 12 show that the results of Multilink's performance follow this result pattern for 3-5 letter words. However, for 6-8 letter words the mean number of cycles start to decline, producing the opposite effect. A first and clear possible explanation for this (unexpected) opposite result pattern for 6-8 letter words lies in the number of words and how the words were used in our simulations. Remember, as told in Section 3.1, that we created our English lexicon by combining the contents of three different databases. The reaction time data we gained from a subset of the English Lexicon Project. While for the graph in Figure 10, the words per category are exactly the same as for Figures 11 and 12, the circumstances under which the reaction times for these words were collected, were different from the circumstances during our simulations.

The lexical decision data in the English Lexicon Project are based on 30 to 35 obser-

vations per item, and each participant providing data for this project, provided data for a subset of approximately 3.000 of the 40.481 tested words. According to New et al. (2006) the number of syllables and the number of orthographic neighbors make independent contributions to the word length effect. While each participant provided data for a subset of the collection of words, he or she had access to his or her background knowledge of the language these words belonged to. We only used a subset of the English Lexicon Project and only provided the words in this subset to Multilink, without access to any extra background knowledge. This means that the number of orthographic neighbors, for example, can never be the same in our simulations as it was when the data were gathered. Therefore, a possible explanation for the decline in our graphs for 6-8 letter words, where for the same words the graph of mean reaction times kept increasing, could be a difference in the number of syllables or the number of orthographic neighbors. By using a subset of the English Lexicon Project, we restricted the number of orthographic neighbors in our lexicon. Having more orthographic neighbors in the set of stimuli leads to more competition, leading to higher number of cycles.

#### 4.4 Multilink's performance for all Dutch words

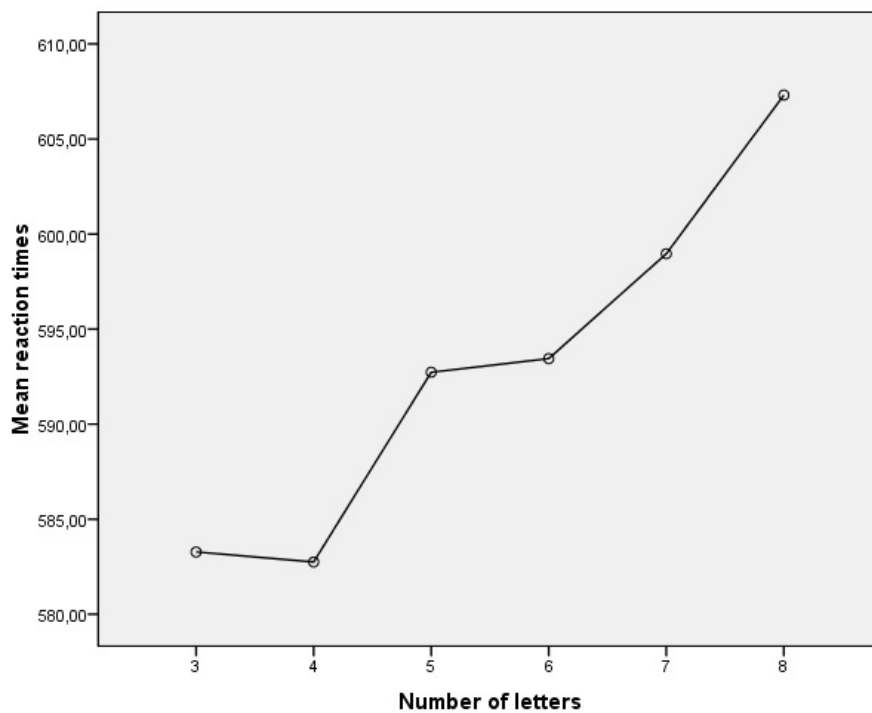
We repeated the same series of simulations, this time for all the Dutch words. The next tables and figures display Multilink's performance on our Dutch lexicon.

	all	3	4	5	6	7	8
OS/SO=0	.504	.566	.584	.515	.571	.577	.536
OS/SO=0.03	.412	.475	.493	.454	.509	.518	.525

**Table 7:** Correlations of the number of cycles in Multilink with reaction times to Dutch words of different lengths. Orthographic-semantic links set to 0 or 0.03.

Table 7 shows that again the correlations computed on the basis of a large number of items (see Table 8) are high. Also on the word length effect the results for the Dutch

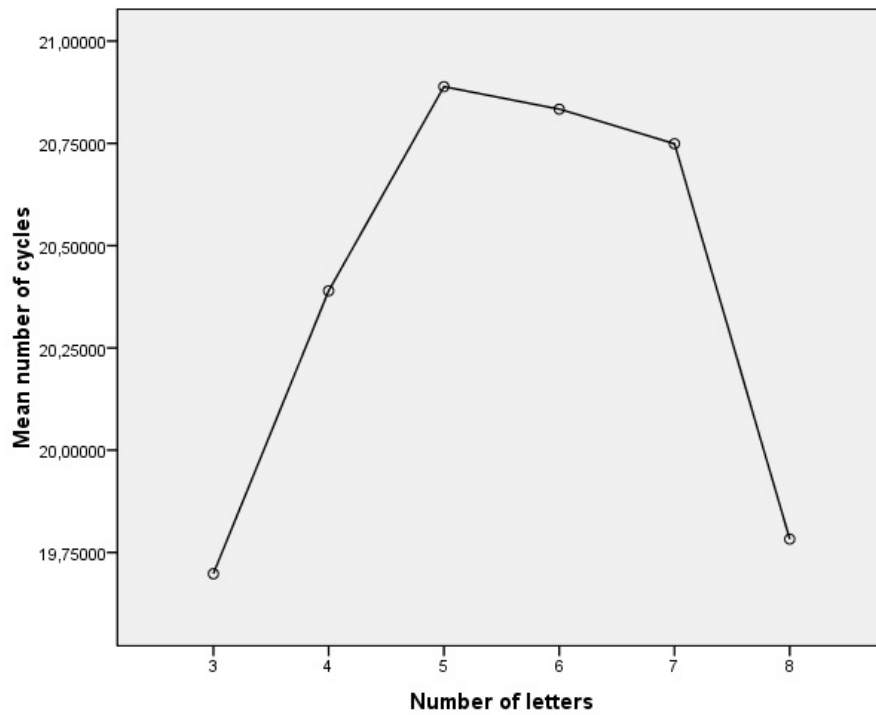
words follow the same pattern as the results for the English words. Again, the word length effect appears to occur in the lexical data of our Dutch lexicon, as shown by Figure 13. Except that this time, the mean reaction time for 4 letter words is slightly lower than that for 3 letter words. Again, our graphs, plotting the mean number of cycles against the number of letters, start to decline for 6-8 letter words. Here the same explanation holds as for the English words, with the addition that the bigger difference between mean cycle times for 7 and 8 letter Dutch words can be explained by the number of words for these categories. Where for the English words, the number of 8 letter words was roughly two thirds the number of 7 letter words, for the Dutch words this is less than half the number of 7 letter words.



**Figure 13:** Word length effect for all our Dutch words, mean reaction times plotted against number of letters

Number of letters	Mean	Standard deviation	N
3	19.70	3.75	182
4	20.39	3.63	470
5	20.89	3.67	503
6	20.83	3.57	541
7	20.75	3.56	335
8	19.78	3.50	149
Total	20.57	3.63	2180

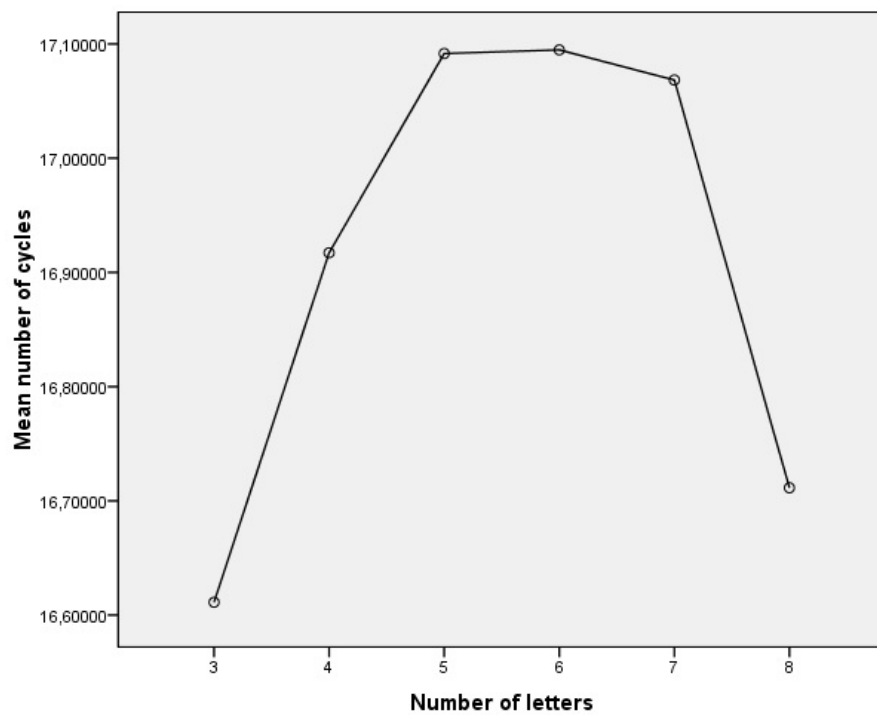
**Table 8:** Mean number of cycles in Multilink per word length for OS/SO=0 for Dutch words.



**Figure 14:** Word length effect Dutch OS/SO=0, mean number of cycles plotted against number of letters

Number of letters	Mean	Standard deviation	N
3	16.61	1.63	182
4	16.92	1.55	470
5	17.09	1.53	503
6	17.09	1.49	541
7	17.07	1.49	335
8	16.71	1.49	149
Total	16.99	1.53	2180

**Table 9:** Mean number of cycles in Multilink per word length for OS/SO=0.03 for Dutch words.



**Figure 15:** Word length effect Dutch OS/SO=0.03, mean number of cycles plotted against number of letters

#### 4.5 Performance of Multilink versus the Bilingual Interactive Activation model on 4 letter words

The results of the previous simulations gave us enough confidence to proceed to the bilingual simulations. We will repeat the previous series of simulations, but this time we will compare Multilink's performance to that of the Bilingual Interactive Activation (BIA) model, which is a bilingual model for word recognition. All the 4 letter words in our bilingual lexicon were fed one by one to Multilink and to the BIA model. Just as in the previous model-to-model comparison, we compared the models with the OS/SO parameter set at 0 and at 0.03. In the previous series of simulations we used Multilink's recognition task. Since we are now exploring the bilingual domain, we switched to Multilink's lexical decision task. To be more specific, we used the standard lexical decision task in which the model decides whether a word belongs to a particular language or not. In total our bilingual Dutch-English lexicon contained 1061 4 letter words, of which 470 were Dutch and 591 were English.

The reason the number of English 4 letter words drops from 850 in the monolingual lexicon to 591 in the bilingual lexicon, is because of the way the lexicons are created and the restrictions on the sets of stimuli (see Section 3.1). The bilingual lexicon contains Dutch-English word pairs. Since we have 2180 Dutch words versus 4017 English words, the bilingual lexicon can also only contain 2180 English words. The bilingual lexicon therefore contains all the Dutch words, including all 470 Dutch 4 letter words, while the number of English 4 letter words has dropped from 850 to 591. The resulting correlations of Multilink and the BIA model with respect to the reaction times, frequency and the logarithm of the frequencies are shown in Table 10.



<b>Dutch N=470</b>	<b>OS/SO = 0</b>			<b>OS/SO = 0.03</b>		
	RTs	Freq	LogFreq	RTs	Freq	LogFreq
Multilink	.522	-.432	-.969	.341	-.454	-.769
BIA	.246	-.206	-.472	.246	-.206	-.472
<b>English N=591</b>	<b>OS/SO = 0</b>			<b>OS/SO = 0.03</b>		
	RTs	Freq	LogFreq	RTs	Freq	LogFreq
Multilink	.371	-.482	-.942	.242	-.435	-.676
BIA	.294	-.319	-.583	.294	-.319	-.583

**Table 10:** Correlations of the number of cycles in Multilink and the BIA model with RTs, Word Frequency, and Log Word Frequency for all four letter words in our Dutch-English lexicon, in a Dutch lexical decision task (top half) and in an English lexical decision task (bottom half).

In Section 4.2 we saw that with the incorporation of semantics, Multilink still outperformed the IA model in the recognition task. From Table 10 we can conclude that with the incorporation of semantics Multilink is also able to outperform the BIA model. Except for the reaction times for the English words when OS/SO is set to 0.03, Multilink performs better than the BIA model on all fronts.

#### 4.6 Multilink's performance for all words in the bilingual lexicon

Just as we investigated how Multilink performs on all the words in our monolingual lexicons, we also investigated how Multilink performs on all the words in our bilingual Dutch-English lexicon. The resulting correlations with reaction times for 2180 English words of different lengths in an English lexical decision task, and 2180 Dutch words of different lengths in a Dutch lexical decision task are shown in Table 11.

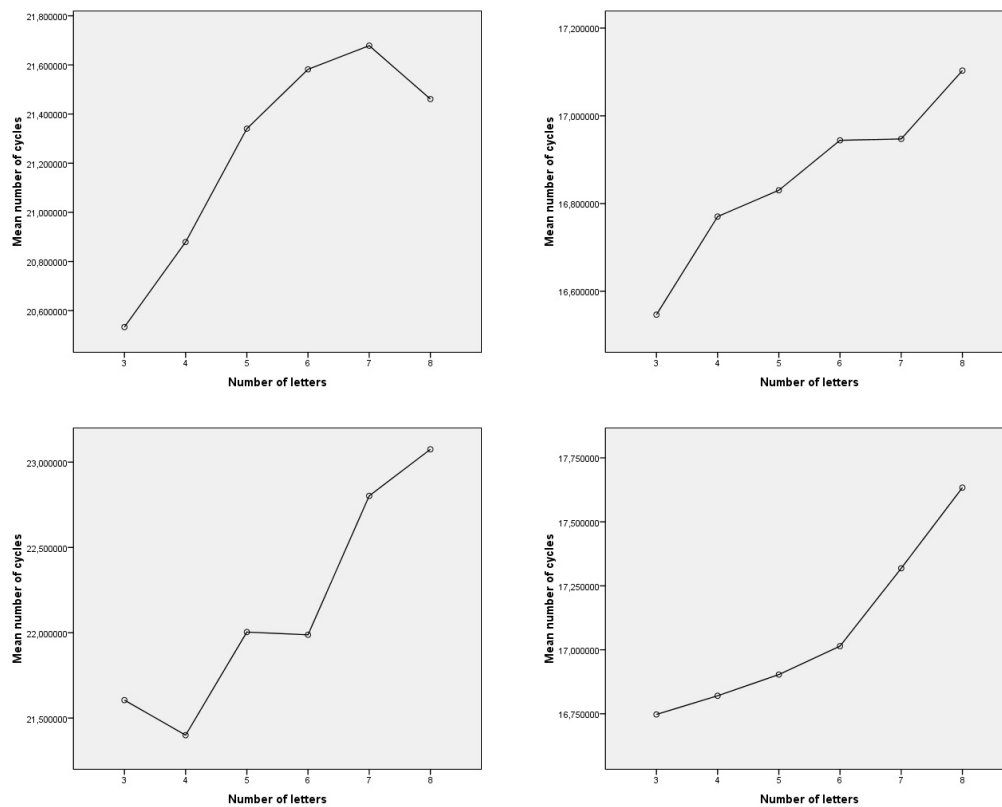
	all	3	4	5	6	7	8
Number of Dutch words	2180	182	470	503	541	335	149
Number of English words	2180	206	591	544	439	265	135

	all	3	4	5	6	7	8
Dutch OS/SO=0	.504	.412	.522	.471	.529	.515	.505
Dutch OS/SO=0.03	.297	.188	.341	.210	.354	.261	.289
English OS/SO=0	.406	.377	.371	.422	.400	.406	.478
English OS/SO=0.03	.272	.266	.242	.258	.248	.316	.385

**Table 11:** Correlations of the number of cycles in Multilink with reaction times to the Dutch and the English words of different lengths in the Dutch-English bilingual lexicon. Orthographic-semantic links set to 0 or 0.03.

Again the correlations of the number of cycles in Multilink with the empirical reaction time data computed on the basis of a large number of items are high. In their paper discussing the BIA model, Dijkstra and Van Heuven (1998) decided not to report correlations for their bilingual simulations on an item basis, because these were all around .15 or lower.

We also performed simulations to determine how well Multilink performs on the word length effect in a bilingual context. For all words in the Dutch-English lexicon of each word length ranging from 3 to 8 letters, we performed an English lexical decision task and a Dutch lexical decision task. The results are shown in the graphs of Figure 16.



**Figure 16:** A graphic overview of the bilingual results on the word length effect. Top-left: English OS/SO=0.00, top-right: English OS/SO=0.03, bottom-left: Dutch OS/SO=0.00, bottom-right: Dutch OS/SO=0.03.

The general trend in each of the graphs is an ascending line, suggesting that it indeed takes longer to reach a lexical decision, the longer the word is. But while these results are based on a smaller set of words (the bilingual lexicon contains 2180 English words and their 2180 Dutch translations), they are more in line with our expectations than the results of the monolingual word length effect simulations. A possible explanation is the different lexicon used for the bilingual simulations. Because we are using a bilingual lexicon we also have a different competitor set as compared to the monolingual word length simulations.

### 4.7 Multilink's performance on cognates and controls

We were also interested to see how Multilink performs on words with cross-linguistic similarities and differences. Therefore we performed simulations to see if Multilink is able to reproduce the translation latencies for cognates and non-cognates as observed in empirical studies. In these simulations we used the non-identical and identical cognates that were used in the studies by Dijkstra et al. (2010) and Vanlangendonck et al. (in preparation).

The study by Dijkstra et al. (2010) was based on 191 English-Dutch word translation pairs, consisting of pairs with perfect orthographic overlap (identical cognates), considerable overlap (non-identical cognates) and control words without much overlap. A 161 of these 191 word pairs also occurred in our bilingual lexicon. Using these 161 word pairs in an English lexical decision task resulted in a correlation (between the number of cycles in Multilink and the reaction times found in the study by Dijkstra et al., 2010) of .32 without semantic links and .34 with semantic links set at 0.03.

For the 167 items in the multilingual lexicon used by Vanlangendonck et al. (in preparation), the correlations between the number of cycles in Multilink (for an English lexical decision task) and the reaction times found in the study by Vanlangendonck et al. (in preparation) were even higher. When semantics was taken into account, the simulation resulted in a correlation of .46, and .37 when semantics were not taken into account. However, it should be noted that for the selected items word frequency and word length were not equally divided across conditions. A closer look at Multilink's performance on the items used by Vanlangendonck et al. (in preparation) reveals that the average cycle times of Multilink follow the same pattern as the average reaction times in the study by Vanlangendonck, as shown in Table 12. As a comparison, for simulations on cognate and false friend processing during L2 learning with other types of models, we refer to Dijkstra et al. (2011).

	EC	IC	LD1	LD2
RTs	647	612	633	634
Multilink	17.33	15.73	16.11	16.84

**Table 12:** Reaction times in the English lexical decision task of Vanlangendonck et al. and number of cycles in Multilink for different item types. EC = English Control word, IC = Identical Cognate, LD1 = Non-identical Cognate with a Levenshtein Distance of 1, LD2 = Non-identical Cognate with a Levenshtein Distance of 2.

#### 4.8 Simulating word translation with Multilink

As described in the introduction, the two research methods most commonly applied in word translation studies are word translation recognition and word translation production. We have reviewed the results of some studies using these research methods and have found that these results are inconclusive. We also implied that the differences in these results give rise to questions about the word translation process which can possibly be clarified by an implemented model of word translation.

By exploring the predictions of Multilink with respect to forward and backward translation production of high and low frequency words in more and less proficient bilinguals, we will test Multilink's performance on the asymmetry predictions made by the RHM. To do this, we will try to answer the following questions:

*Question 1:* Does the model predict an asymmetry in forward and backward translation? As mentioned earlier, the results with regard to this aspect of word translation are not conclusive. Where Kroll et al. (2002) found backward translation to be faster, Christoffels et al. (2006) found forward translation to be faster.

*Question 2:* Is this asymmetry in response times sensitive to L2 proficiency and word frequency? Remember that Christoffels et al. (2006) found a larger asymmetry for less proficient bilinguals (students) than for more proficient bilinguals (teachers and interpreters). The same was true for Kroll et al. (2002), but in the opposite direction.

---

The bilingual simulations performed in order to answer these questions, were realized in collaboration with Steven Rekké.

Christoffels et al. also reported that high-frequency words were translated faster than low-frequency words.

*Question 3:* Is there a cognate facilitation effect in translation, as reported by most studies, e.g. Christoffels et al. (2006)? If so, can Multilink clarify whether this effect comes from comprehension and/or production?

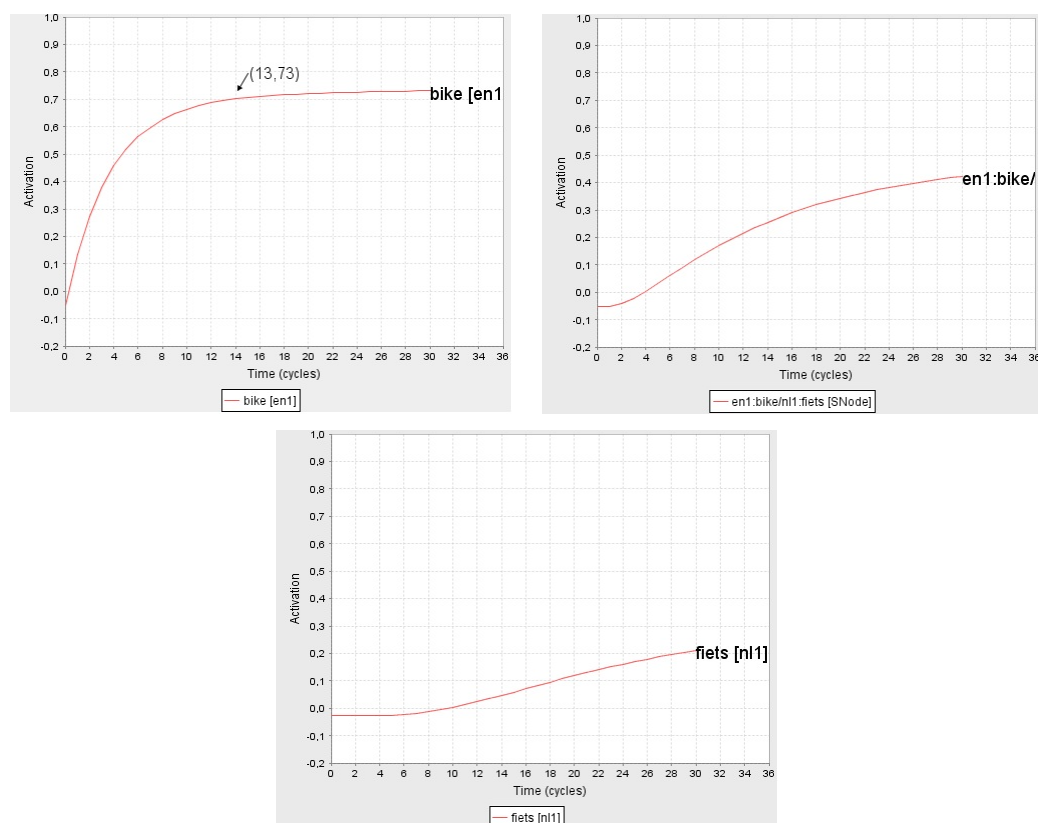
*Question 4:* Which of the three result patterns of non-cognate translations in Table 13 (the studies of particular interest to us in regard to the assumptions made in the RHM, see Page 16) does Multilink produce? First, is it that of Christoffels et al. (2006), a larger asymmetry for low proficient bilinguals than for high proficient bilinguals, with for low proficient bilinguals significantly faster forward ( $L1 \rightarrow L2$ ) translation than backward ( $L2 \rightarrow L1$ ) translation, but not for high proficient bilinguals, they translate almost just as fast in both directions? Second, is it that of De Groot et al. (1994) and De Groot (1992), who reported faster backward translation than forward translation, or just as fast, for low proficient bilinguals? Third, is it the pattern of Kroll et al. (2002), who just as Christoffels et al. (2006) found a larger asymmetry for low proficient bilinguals than for high proficient bilinguals, but in the opposite direction? They found faster backward than forward translation.

Christoffels et al. (2006)						
	HP bilingual			LP bilingual		
	HF	LF	All	HF	LF	All
L1 → L2	782	1083	932	883	1149	1016
L2 → L1	803	1055	929	959	1225	1092
	3			-78		
De Groot et al. (1994) and De Groot (1992)						
	LP			LP		
L1 → L2	1463			1315		
L2 → L1	1390			1328		
	73			-13		
Kroll et al. (2002)						
	HP			LP		
L1 → L2	1104			1511		
L2 → L1	964			1223		
	140			288		

**Table 13:** Response times to non-cognates in three studies involving forward and backward translation.

HP = High Proficient (interpreters); LP = Low Proficient (students); HF = High Frequency items;  
LF = Low Frequency items.

The word translation process in Multilink goes from orthography, to semantics, to phonology. First a letter string presented to the model as input activates orthographic input representations. Then, based on these input representations, semantic representations are activated (comprehension). These in turn activate a phonological output string (production). This process is visualised in Figure 17. As mentioned in Section 1.4, we have not used the word association route in our version of Multilink, meaning that the weight on L1 - L2 word form links are set to zero.



**Figure 17:** A graphic example of the word translation process: From English orthographic input BIKE (top-left panel) via semantics (top-right panel) to Dutch phonological output FIETS (bottom panel).

We performed several word translation simulations with Multilink for 550 non-cognates with a Levenshtein distance larger than 2. The results of these simulations are presented in Table 14. In the simulations for high proficient bilinguals, the frequencies as recorded in the CELEX database were used. For the low proficient bilingual simulations, the English frequencies have been divided by 4, and the resting levels have been adapted in such a way that they do not influence the resting level activations of the Dutch words (see Section 3.2). Another difference between the simulations for high and low proficient bilinguals is the strength of the links between orthography and semantics, or semantics and phonology. In determining these strengths we made the following assumptions; high proficient bilinguals have about as much experience trans-



lating in forward direction as they do in backward direction; low proficient bilinguals translate slower than high proficient bilinguals; and low proficient bilinguals are proficient in L1 but not in L2. Based on these assumptions, and previous trial-and-error simulations to determine the best parameter settings in a translation simulation, we used the following weight settings in our translation simulations to simulate different levels of proficiency: for high proficient bilinguals  $O \rightarrow S = .03$ , and  $S \rightarrow P = .3$  (for both translation directions). For low proficient bilinguals in forward translation  $O \rightarrow S = .03$ ,  $S \rightarrow P = .25$ , and in backward translation  $O \rightarrow S = .025$ ,  $S \rightarrow P = .3$ .

Multilink						
	HP bilingual			LP bilingual		
	HF	LF	All	HF	LF	All
L1 $\rightarrow$ L2	26.50	27.40	26.95	32.70	33.41	33.05
L2 $\rightarrow$ L1	26.49	27.50	26.99	28.49	29.17	28.83
	26.97			30.94		

**Table 14:** Multilink's number of cycles for non-cognates in forward and backward translation between Dutch and English. HP = High Proficient; LP = Low Proficient; HF = High Frequency items; LF = Low Frequency items.

Comparing Table 14 to Table 13 answers our 4 questions as follows.

*Question 1:* The number of cycles in the All column for balanced (HP) bilinguals are almost equal, whereas they differ a few cycles for unbalanced (LP) bilinguals. This implies that Multilink does indeed predict an asymmetry in forward and backward translation, but only for unbalanced bilinguals. Just as reported by Kroll et al. (2002), for these bilinguals the translation from L1 to L2 is slower than from L2 to L1. Forward and backward translation for balanced bilinguals are symmetrical, as reported by Christoffels et al. (2006). We can interpret this as the mutual effect between words of two languages being more or less the same, when L1 and L2 are equally strong (which may be expected in professional interpreters). The asymmetry for low proficient bilinguals is an indication that L2 output words profit more from L1 input words than L1

output words from L2 input words. This is because a strong input boosts lexical activation and it takes the translation less activation afterwards to reach the recognition threshold.

*Question 2:* Just as Christoffels et al. (2006) and Kroll et al. (2002) found, Table 14 demonstrates that the size of the asymmetry in translation production in Multilink is sensitive to L2 proficiency. It is also sensitive to word frequency, just like Christoffels et al. (2006) predicted. In Multilink, the speed of lexical activation spreading through the network depends on two factors. The first of this is the resting level activations, which for L2 are generally lower in unbalanced bilinguals, and are lower for low-frequency input words. The second factor is the strength of the connections between semantics and word forms (both orthographic and phonological word forms), which are also generally weaker in unbalanced bilinguals for L2.

What is most interesting about the asymmetry predictions made by Multilink, is that Multilink is able to reproduce the asymmetry predictions made by the RHM (the presence of an asymmetry effect in the direction of faster backward translation for low proficient bilinguals), without the involvement of the word association route. Where the RHM suggests that this effect is due to word form links between L1 and L2 word form representations, these word form links are not present in our version of Multilink. This means that the assumption made in the RHM is not supported by Multilink, which is a clear example of how an implemented computational model can be used to prove or disprove assumptions made in a theoretical model.

*Question 3:* At the time of writing we were not yet able to optimise the parameters for identical and non-identical cognates. The number of cycles for cognates are therefore absent from Table 14. For now, we only ran preliminary simulations in order to try and answer this question. They show the presence of a very strong cognate facilitation effect (in the order of a difference of 4 cycles). This cognate facilitation effect in Multilink can be understood as follows. The input letter string does not only activate the cognate reading in the input language, but also coactivates the cognate reading in the output language. The orthographic reading of this output cognate reading activates its phonol-

ogy, resulting in a direct and extremely strong cognate effect. The consequences of this simulation finding for word translation and for Multilink are considered in Section 5.

*Question 4:* Table 14 shows that the result patterns of Multilink for non-cognates are a combination of those in Christoffels et al. (2006) and Kroll et al. (2002), namely that there is a larger asymmetry for low proficient bilinguals than for high proficient bilinguals, and high proficient bilinguals translate almost just as fast in both directions (like in Christoffels et al., 2006), but low proficient bilinguals translate faster in backward than in forward translation (like in Kroll et al., 2002). We define non-cognates as pairs of translation equivalents with limited form overlap (e.g., *vogel* - *bird*, *bike* - *fiets*). Because for non-cognates the input and output representations do not overlap, there is no direct link between representations used for comprehension and production, in contrast to cognates.

## 5 Conclusions and discussion

In this thesis we have investigated and developed Multilink, a computational model for the domain of bilingual word recognition and word translation introduced by Dijkstra and Rekké (2010). The general architecture of the model is based on the localist-connectionist paradigm, with the addition of some new features, like the specification of orthographic differences between lexical items in terms of Levenshtein distance.

The most prominent advantage of a computational model compared to a verbal model is the requirement that every assumption must be made explicit. Implementing a theory into a computer model, forces one to formalize the theory into a set of algorithms such that a machine is able to execute it. A successfully implemented model allows for fast and accurate quantitative and qualitative simulations. It thereby provides the opportunity to test the underlying theory on up-to-date empirical data, unlike verbal models, such as the RHM, which are often underspecified and incomplete.

A verbal model can lead to interpretation problems with respect to model predictions. And being only a verbal model makes it difficult to test whether all assumptions

made by the model are critical for its processing or not. In the RHM for instance, it has been relatively hard to confirm or reject the functionality of the word association route.

Having all the underlying assumptions in Multilink made explicit brings with it the advantage that it helps us to better understand the word translation process. It gives us insight in how multiple underlying mechanisms interact and play out. We can add or leave out particular assumptions to test whether or not they are critical in achieving a certain reaction time pattern. This gives the model heuristic value, because it allows simulations of complex interactions between multiple variables.

In an earlier paper Dijkstra and Rekké (2010) presented a beta-version of Multilink. It provided some simulations for a limited number of words to test some of the assumptions at face value. The main research question in this thesis was therefore: *Can we validate Multilink based on empirical data?* To answer this question we have introduced some additions to Multilink, we greatly extended the lexicons used by Multilink and we introduced a way to simulate different levels of L2 proficiency.

We first tried to validate Multilink in the monolingual domain by answering two questions: *How does Multilink perform compared to a computational model for monolingual word recognition, the Interactive Activation (IA) model?* and *Can Multilink reproduce the empirical effects found in the psycholinguistic domain?* In our monolingual simulations we found remarkably high correlations between Multilink and empirical data in a model-to-data comparison. A model-to-model comparison of Multilink to the IA model yielded a favorable outcome for Multilink. Only on the word length effect, did the model not fully meet our expectations. At the end of our research a paper by New et al. (2006) came to our attention, which not only provides a thorough summary of earlier work on the word length effect, but also describes a series of analysis on this effect based on words from the English Lexicon Project. Due to time constraints we could only use the information in this paper for the evaluation of our results on the word length effect. However, the work done by New et al. does provide a very interesting opportunity for future research with Multilink.

New et al. (2006) have shown that there are several factors, e.g. number of syllables

and number of orthographic neighbors, that all make independent contributions to the word length effect. The choices we made in the creation of our datasets did not allow us to control for these factors, at least not to the extent that New et al. (2006) did. They also discuss implications for models of visual word recognition. Although visual word recognition is not the main purpose of our model, it is a part of the whole that is Multilink. When this part of Multilink was designed, we did not yet have knowledge of these suggested implications. Therefore it would be very interesting to see how well Multilink would perform on tasks used by New et al. (2006).

Overall, Multilink's results of the monolingual simulations, in both the model-to-model comparison as well as the model-to-data comparisons, provide a solid basis for its extension to bilingual processing.

In the bilingual domain, we again performed model-to-model and model-to-data comparisons, to answer the question: *How well does Multilink perform compared to the Bilingual Interactive Activation (BIA) model and can Multilink also reproduce the empirical effects in bilingual simulations?* Just as with the comparison to the IA model, this comparison yielded a favorable outcome for Multilink. Again, this time in the bilingual domain, Multilink was able to produce high correlations with empirical data. We also re-examined the word length effect in the bilingual domain, of which the outcome was in line with our expectations.

Furthermore, we applied Multilink to investigate two questions with respect to bilingualism. First, *How will Multilink perform on the general asymmetry predictions for forward and backward word translation made by the RHM?* Second, *Can Multilink reproduce the translation latencies for cognates and non-cognates as observed in empirical studies?*

Multilink is partly built with the RHM in mind and can therefore be considered as a partial implementation of RHM. As such, we were able to test the asymmetry predictions made by RHM in a quantitative way (but without word association involvement). In our simulations, an asymmetry with respect to translation directions was observed for low proficient bilinguals. We found that L1 words were translated slower into L2 words than vice versa, which is in line with Kroll et al. (2002). Kroll and Stewart (1994)

believed that this asymmetry is the result of lexical links between L1 word form representation and L2 word form representations as depicted in Figure 5.

The lexical link from L2 to L1 (in Figure 5) is stronger than the link from L1 to L2, to reflect the way L2 was learned. Kroll and Stewart believe that during the acquisition of L2, bilinguals learn to associate every word form from L2 with its equivalent in L1. Thereby forming a strong lexical-level association. However, the same asymmetry was observed in Multilink’s simulations, without any involvement of the word association route. Therefore, the assumption made by the RHM that especially low proficient bilinguals profit from word association is not supported.

Our bilingual Multilink simulations with respect to cognate and non-cognate processing had a considerable heuristic value and also led to some fortunate findings. First, although not reported, Multilink produced a very large cognate facilitation effect in translation (see Question 3, Page 49). The effect in comprehension (Section 4.7, where we used a word recognition task) was more realistic. The orthographic overlap of the two readings of a cognate is the cause of this finding. The orthographic input of a cognate partially activates its phonological output representation. This causes a direct and extremely large cognate facilitation effect. As such, this mechanism represents a fourth source for cognate facilitation in language production, in addition to the three specified by De Groot (2011, p.262). This problem would not have existed when completely separate phonological representations would have been used for comprehension and production. But since the model is assumed to be completely interactive, it is hard to imagine that no “leakage” would take place all. What we can learn from this fortunate finding is that the output for word translation is probably not selected on the basis of phonological activation only.

Some models, like the WEAVER++ model, include language membership in their output selection. They include the requirement that the output item must belong to the target language. We could add this requirement to our phonological activation criterion, but this addition alone would still not be sufficient. A low-frequency L2 input letter string may activate a high-frequency competitor from the L1 output language.

This competitor then becomes phonologically activated and it belongs to the target language, yet it is not the word we were looking for as output.

To overcome this problem we propose the following solution for future research. An activated phonological output item that belongs to the right language will only be accepted as a correct translation if there is enough activation of the associated semantic representation. This reduces the influence of orthographic overlap on phonological activation and implies that the link between phonology and semantics is checked before the word translation is selected and produced.

A second consequence of our simulation work is that without making further assumptions, word association will not work as a translation route. First, it aggravates the cognate problem we just discussed, because the competitor is also activated directly by the form link to the input reading, and is not only coactivated together with the input. Second, because there are other word associations as well (e.g., in fixed expressions), the associated word must still be checked as a correct translation. Third, word translations can have multiple mappings (Tokowicz and Kroll, 2007). This implies that several possible translations of the input word will become activated as possible translations. When considering all these possible translations in isolation, no selection is possible, except when additional criteria (like frequency of mapping) are used. However, in sentence translation, the selection of the right translation depends on the sentence-context. This selection requires the use of semantics, not only that of the rest of the sentence, but also that of the item set on which the selection occurs. This is therefore yet another reason that the semantic link between input and output must standardly be checked.

The applications of the model discussed in this thesis are only part of the potential that Multilink has in the field of psycholinguistics. The model has already been equipped to do much more: e.g., semantic priming (using word association information), false friend processing and picture naming. Also, in the scope of this thesis, we have not spent research time on task-dependent processing. These issues, as well as our suggestions for future research on the word length effect with Multilink, will be considered one by one in future studies.

To sum up, the main goal of this thesis was to validate Multilink based on empirical data. In both the monolingual and the bilingual simulations, we saw that Multilink produces a high correlational fit to empirical data. Further, we saw that in both model-to-model comparisons Multilink was the better performing model. We have proven that Multilink is able to reproduce the translation latencies for cognates and non-cognates in a word recognition task. In a word translation task we have proven that Multilink is able to reproduce the translation latencies for non-cognates, but not for cognates. Furthermore, in our translation simulations we were able to demonstrate that Multilink can reproduce the asymmetry predictions for forward and backward word translation made by the RHM, but that it can do so without the involvement of word association. These results provide more than sufficient validation for Multilink to stimulate future research with the model.

## References

- Altaribba, J. and Mathis, K. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language*, 36:550–568.
- Baayen, H., Piepenbrock, R., and Van Rijn, H. (1993). The CELEX lexical database. CD-ROM. Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium.
- Balota, D. (1994). Visual word recognition: The journey from features to meaning. In Gernsbacher, M., editor, *Handbook of Psycholinguistics.*, pages 303–358. Academic Press, San Diego.
- Balota, D., Yap, M., Cortese, M., Hutchison, K., Kessler, B., Loftis, B., and Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, 39:445–459.
- Bloem, I. and La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for model of lexical access in language production. *Journal of Memory and Language*, 48:468–488.



- Bloem, I., Van Den Boogaard, S., and La Heij, W. (2004). Semantic facilitation and semantic interference in language production: Further evidence for the conceptual selection model of lexical access. *Journal of Memory and Language*, 51:307–323.
- Brenders, P., Van Hell, J., and Dijkstra, T. The development of lexical and conceptual processing in L2 learners: Electrophysiological and behavioral studies. Under revision.
- Christoffels, I., De Groot, A., and Kroll, J. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, 54:324–345.
- Comesaña, M., Perea, M., Piñeiro, A., and Fraga, I. (2009). Vocabulary teaching strategies and conceptual representations of words in L2 in children: Evidence with novice learners. *Journal of Experimental Child Psychology*, 104:22–33.
- De Deijne, S. and Storms, G. (2008). Word association: Norms for 1424 Dutch words in a continuous task. *Behavior Research Methods*, 40:198–205.
- De Groot, A. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18(5):1001–1018.
- De Groot, A., Dannenburg, L., and Van Hell, J. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language*, 33:600–629.
- De Groot, A. and Poot, R. (1997). Word translation at three levels of proficiency in a second language: The ubiquitous involvement of conceptual memory. *Language Learning*, 47:215–264.
- Dijkstra, A., Haga, F., Bijsterveld, A., and Sprinkhuizen-Kuyper, I. (2011). Assessing mechanisms of inhibition in localist and distributed connectionist models by simulating L2 word acquisition. In Altarriba, J. and Isurin, L., editors, *Memory and Language: Theoretical and Applied Approaches to Bilingualism*. John Benjamins, Hillsdale, NJ. In press.

- Dijkstra, A., Lormans, E., Rekké, S., and Roelofs, A. (2012). Multilink: a computational model for word translation. *Bilingualism: Language and Cognition*. Under review.
- Dijkstra, A., Miwa, K., Brummelhuis, B., Sappelli, M., and Baayen, H. (2010). How cross-language similarity affects cognate recognition. *Journal of Memory and Language*, 62:284–301.
- Dijkstra, A. and Rekké, S. (2010). Towards a localist-connectionist model for word translation. *The Mental Lexicon*, 5(3):403–422.
- 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. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Dijkstra, A. and Van Heuven, W. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5:175–197.
- Duijck, W. and Brysbaert, M. (2004). Forward and backward number translation requires conceptual mediation in both balanced and unbalanced bilinguals. *Journal of Experimental Psychology: Human Perception and Performance*, 30:889–906.
- Dyer, F. (1973). The stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition*, 1:106–120.
- Ferré, P., Sánchez Casas, R., and Guasch, M. (2006). Can a horse be a donkey? Semantic and form interference effects in translation recognition in early and late proficient and non-proficient Spanish-Catalan bilinguals. *Language Learning*, 65(4):571–608.
- Francis, W. and Gallard, S. (2005). Concept mediation in trilingual translation: Evidence from response time and repetition priming patterns. *Psychonomic Bulletin & Review*, 12:1082–1088.
- Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes*, 23:1–35.

- Keuleers, E., Diependaele, K., and Brysbaert, M. (2010). Practice effects in largescale visual word recognition studies: a lexical decision study on 14000 dutch mono- and disyllabic words and nonwords. *Frontiers in Language Sciences: Psychology*, 1:174.
- Kroll, J., Micheal, E., Tokowicz, N., and Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, 18:137–171.
- Kroll, J. and Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33:149–174.
- La Heij, W., Hooglander, A., Kerling, R., and Van Der Velden, E. (1996). Nonverbal context effects in forward and backward word translation: Evidence for concept mediation. *Journal of Memory and Language*, 35:648–665.
- Linguistic Systems B.V. (2008). Euroglot Professional 5.0. <http://www.euroglot.nl/>.
- McClelland, J. and Rumelhart, D. (1981). An interactive activation model of context effects in letter perception, Part 1: An account of basic findings. *Psychological Review*, 88:375–405.
- Miller, N. and Kroll, J. (2002). Stroop effects in bilingual translation. *Memory & Translation*, 30:614–628.
- Nelson, D., McEvoy, C., and Schreiber, T. (1998). The University of South Florida word association, rhyme and word fragment norms. <http://www.usf.edu/FreeAssociation/>.
- New, B., Ferrand, L., Pallier, C., and Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, 13:71–78.
- Potter, M., So, K., Von Eckardt, B., and Feldman, L. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 23:23–38.

- Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, 64:249–284.
- Roelofs, A. (2002). How do bilinguals control their use of languages? *Bilingualism: Language and Cognition*, 5:214–215.
- Schepens, J., Dijkstra, A., and Grootjen, F. (2011). Distribution of cognates in Europe as based on Levenshtein distance. *Bilingualism: Language and Cognition*. In press.
- Sholl, A., Sankaranarayanan, A., and Kroll, J. (1995). Transfer between picture naming and translation: A test of asymmetries in bilingual memory. *Psychological Science*, 6:45–49.
- Sunderman, G. and Kroll, J. (2006). First language activation during second language lexical processing. *SSLA*, 28:387–422.
- Talamas, A., Kroll, J., and Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition*, 2:45–58.
- Tokowicz, N. and Kroll, J. (2007). Number of meanings and concreteness: Consequences of ambiguity within and across languages. *Language and Cognitive Processes*, 22:727–779.
- Tokowicz, N., Kroll, J., De Groot, A., and Van Hell, J. (2002). Number-of-translation norms for Dutch-English translation pairs: A new tool for examining language production. *Behavior Research Methods, Instruments and Computers*, 34:435–451.
- Van Heuven, W., Dijkstra, A., and Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, 39:458–483.
- Vanlangendonck, F., Rueschemller, S., and Dijkstra, T. Cross-linguistic similarity effects for cognates and false friends: behavioral and fMRI evidence. [working title]. In preparation.