

# Multilink: a model for multilingual processing

Steven T. Rekké

Department of Artificial Intelligence

Radboud University Nijmegen

Correspondence: [steven@rekke.net](mailto:steven@rekke.net)

Artificial Intelligence BSc. Thesis

Version: Februari 17, 2010 Supervisor: Prof. dr. Ton Dijkstra

Donders Centre for Cognition

Radboud University Nijmegen

**Abstract**

In this paper, a new model of multilingual processing, called Multilink model, is developed that can account for cognate processing, bilingual semantic priming, and word translation. To provide a theoretical background, principles of computational modeling are first discussed, followed by a consideration of influential psycholinguistic models for word perception and production. As a first exploration of the model's capabilities and properties, simulation studies of basic psycholinguistic phenomena (such as word frequency, word length, and semantic priming effects) are presented. Next, a comparison is made of simulation results to actual empirical data with respect to lexical decision and language decision. In the general discussion, the model's performance is evaluated. A positive aspect of the model is its capability of processing words from different languages and of different lengths and of simulating the translation of words across languages. Some possibilities for future research are also considered.

---

Special thanks to Ton Dijkstra for his guidance, valuable advice, never fading enthusiasm and patience, and to all other people who have supported me along the way.

## Contents

<b>Introduction</b>	<b>6</b>
Motivations . . . . .	7
Influential models . . . . .	8
The Interactive Activation model . . . . .	8
The Bilingual Interactive Activation model . . . . .	11
The BIA+ model . . . . .	12
The Revised Hierarchical Model . . . . .	14
WEAVER++ . . . . .	15
Research questions . . . . .	15
Expectations . . . . .	17
<b>Multilink: a new model for multilingual processing</b>	<b>18</b>
General description . . . . .	19
Structural description . . . . .	19
Word Identification System . . . . .	20
Task/Decision System . . . . .	25
Processing description . . . . .	29
Input . . . . .	29
Word identification . . . . .	30
Translation . . . . .	31
<b>Preliminary simulation studies</b>	<b>32</b>
Simulation study 1: Word frequency . . . . .	32
Goal . . . . .	32
Method . . . . .	32
Results . . . . .	33
Discussion . . . . .	34

Simulation study 2: Word length . . . . .	34
Goal . . . . .	34
Method . . . . .	34
Results . . . . .	35
Discussion . . . . .	35
Simulation study 3: Semantic priming . . . . .	36
Goal . . . . .	37
Method . . . . .	37
Results . . . . .	37
Discussion . . . . .	39
<b>Empirical Data Studies</b>	<b>41</b>
Goal . . . . .	41
Method . . . . .	41
Stimuli . . . . .	43
Parameter settings . . . . .	43
Results . . . . .	43
Language Decision . . . . .	43
Lexical Decision . . . . .	47
Discussion . . . . .	47
<b>General Discussion</b>	<b>49</b>
<b>References</b>	<b>52</b>
<b>Default parameters</b>	<b>54</b>
<b>List of words used in simulations</b>	<b>54</b>
Simulation study 1 . . . . .	54
Simulation study 2 . . . . .	54

<i>MULTILINK</i>	5
Simulation study 3 . . . . .	54
Empirical data study 1 . . . . .	54
Empirical data study 1 . . . . .	54
<b>Calculation of reaction time distributions</b>	<b>54</b>

## Introduction

Communication through spoken and written language is one of the most basic human skills. For decades, even centuries, researchers in the field of psycholinguistics have strived at better understanding the processes and mechanisms underlying language comprehension and production. The vast majority of this research has been performed in the monolingual domain. This is remarkable, because according to Romaine (1995), there are about thirty times as many languages (about 6,000) as there are countries (about 200), and, traditionally, the use of several languages in one country has been rule rather than exception (Grosjean, 1982).

What is even more remarkable is that we are able to acquire and use several languages more or less simultaneously, and that we can translate words and sentences from one language to another. This human capacity is of vital importance to economic, cultural, and scientific development in regions where multiple languages are spoken. The European Union (EU) for instance, with its 27 different countries, relies heavily on people being able to use and translate multiple languages (see Eurobarometer 63.4, 2005).

People who can hold at least a simple conversation in more than one language can be defined as bilinguals (for two languages) or multilinguals (for two or more languages) (Grosjean, 1981). By convention, the primary language (or mother tongue) of a bilingual is referred to as ‘L1’, while the second language is referred to as ‘L2’. We know from literature in the field that translation is quite a complex behavior, because when a bilingual perceives a word it “activates” word possibilities in both known 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 this behavior is very complex, it is available to all bilinguals, including, for instance, immigrant children translating for their parents. How do bilinguals and multilinguals perform this complex feat so easily and efficiently?

*Motivations*

Developments in communication technologies (like the World Wide Web and its dozens of online communities or Instant Messenger services) over the years are stimulating the use of another language next to the mother language even further. As it is becoming easier to meet people from other countries ‘online’, the need to be able to communicate in other languages is likely to increase as well. Because the number of multilinguals in the world is not likely to decrease in the near future, the phenomenon of bilingualism deserves serious research attention.

While the number of empirical studies in the multilingual domain, in particular with respect to translation, is already limited, studies involving a modeling approach are even scarcer. Nevertheless, modeling is an important approach to research, because it allows researchers to formalize their ideas, eliminate ambiguities and check the underlying theory for consistency. In addition, models can inspire the modeler to come up with new ideas. Implemented models are especially useful because they force the modeler to formalize a theory in terms of a set of algorithms to a level where a machine is able to execute it. The modeler has to be explicit about all aspects of the model, and the major advantage to this is that no assumptions remain implicit or hidden. A less fortunate consequence is that modelers often need to make less well motivated assumptions, but at least they will be explicit. A computational model embodies a set of coherent, testable constraints (assumptions) on the relation between input and output. Because such a model allows for fast and accurate quantitative and qualitative simulations, including simulations of complex interactions between many different variables, it gives us the opportunity to test assumptions arising from the underlying theory on actual and up-to-date empirical data.

As mentioned above, translation links word recognition in one language to word production in another language. Implemented computer models exist for both components, but no implemented model currently exists that combines the two into a model of the whole translation process. The Revised Hierarchical Model (RHM) (Kroll & Stewart, 1994) provides a verbal account, but is not implemented. The Bilingual Interactive Activation (+)

model (BIA / BIA+) (van Heuven, Dijkstra, & Grainger, 1998) (Dijkstra & van Heuven, 2002a) (Dijkstra & van Heuven, 2002b) only simulates the word recognition process in proficient bilinguals and in monolinguals. The WEAVER++ model (Roelofs, 1997), a hybrid computer model, describes (especially) the monolingual process of producing spoken word forms on the basis of presented pictures or word forms. In this study we propose a new computational model for multilingual processing, including translation, and compare it to up-to-date empirical evidence.

### *Influential models*

In the following, we will first discuss some of the models that have inspired our model. The first model to be discussed is a bilingual variant of a model that resulted from research on visual word recognition in the monolingual domain, namely the Interactive Activation model.

#### *The Interactive Activation model*

The Interactive Activation model of visual word recognition (or IA model) by McClelland and Rumelhart (1981) is one of the most influential computational models of word recognition to date. The IA model and its extension to the bilingual domain by van Heuven et al. (1998), have inspired the model that will be later proposed. The IA model is a localist-connectionist model, implying that the symbolic units in the model represent ‘objects in the world’ and that these representations are local (i.e. one unit per ‘object’) as opposed to distributed. Furthermore, it implies that these units are connected to each other through excitatory or inhibitory connections. In short, the model consists of multiple levels of symbolic representations connected in a network of nodes. The nodes in the bottom level in this network correspond to the visual features of letters, e.g., a vertical line. The nodes in the level above correspond to the letters themselves, e.g., ‘B’. The nodes in the top layer correspond to the full orthographic (written form) representations of words, e.g., ‘BIKE’. The letter and feature layer are referred to as the sub-lexical layers of the network.

The structure of the IA model is shown in Figure 1. When visual input is presented



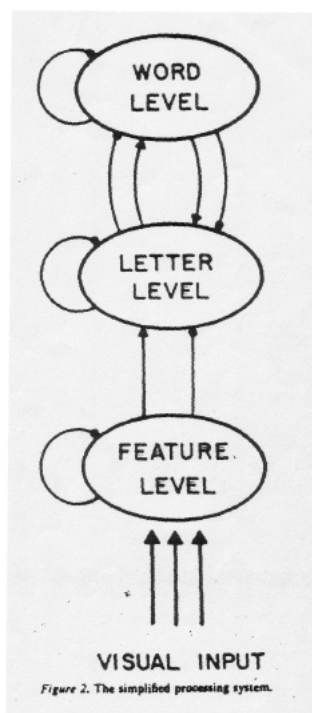


Figure 1. The Interactive Activation model

to the model, nodes on the feature level representing the visual features in the input become activated. They then spread their activation to the letters in which they occur and inhibit letters in which they do not occur. These letters in turn spread their activation to the words in which they occur and inhibit words in which they do not. From the words there is top-down feedback of activation back to the letter layer, so words activate the letters that are present in that word and inhibit the ones that are not. On all layers there is a possibility for competition between nodes via a process, called lateral inhibition, where nodes on the same layer inhibit each other. In fact, lateral inhibition on the letter and feature layers is only used by the authors in early simulation studies. For most purposes, only lateral inhibition on the word level is allowed. Finally, a word is recognized if its activation surpasses some threshold, usually at the value of 0.7.

The IA model can account for many effects known in the field of word recognition (McClelland & Rumelhart, 1981). Unfortunately, it is not capable of dealing with multiple

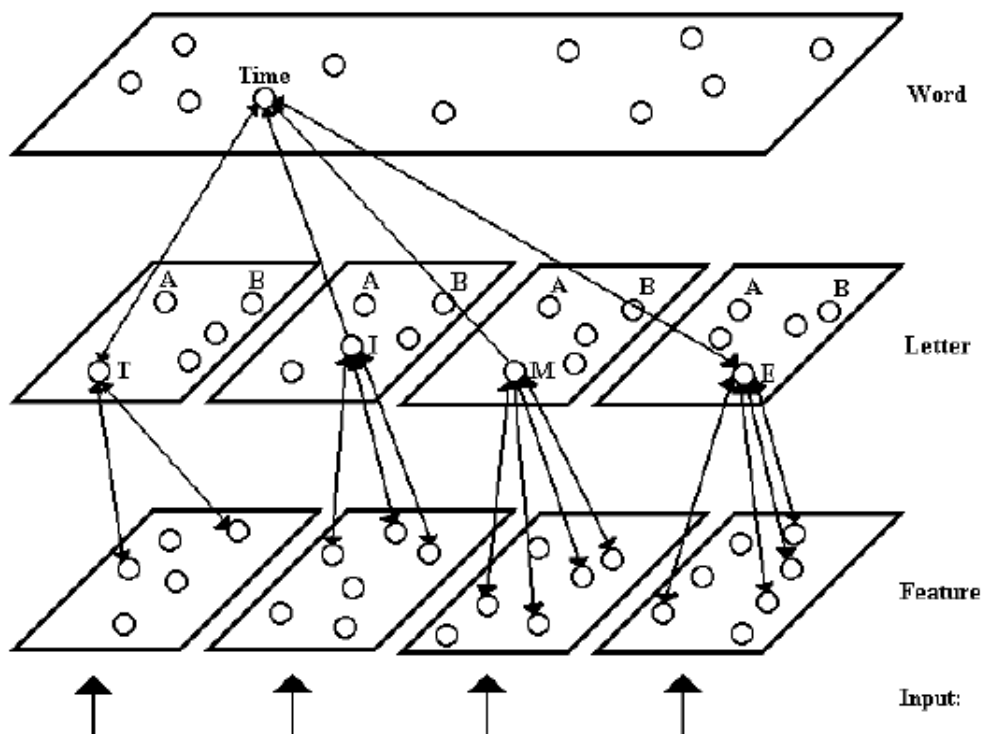


Figure 2. Symbolic representation at different levels of the IA model

languages and it does not incorporate a word's semantic representation (its meaning) or phonological representation (its spoken form). Note that the letters in a word are position-specifically encoded. This is shown more clearly in Figure 2. Position-specific encoding is necessary for the model to be able to distinguish the word 'RAT', for instance, from the word 'TAR' (otherwise both words would receive equal activation upon the input of 'RAT'). The position-specific coding has one main disadvantage, namely that it restricts word recognition to words of the same length (e.g., the model does not pick up the similarity of the words 'MOTHER' and 'OTHER'). Usually, only four letter words are used. An example of part of the network after giving the model the letter 'T' as input for the first letter position is shown in Figure 3.

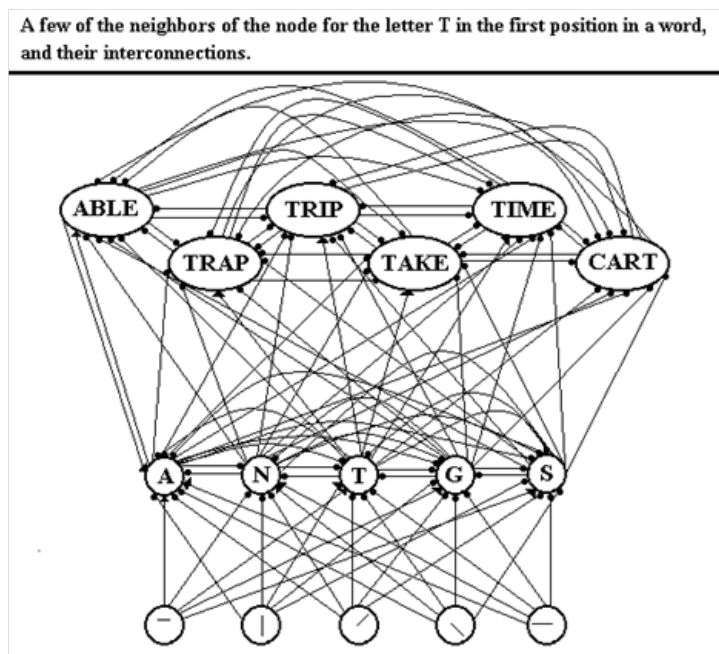


Figure 3. The part of the network in the Interactive Activation model that is relevant for recognizing words with the letter T at the first position.

#### *The Bilingual Interactive Activation model*

Using a nested modeling approach, van Heuven et al. (1998) proposed the Bilingual Interactive Activation (BIA) model and thereby extended the IA model to the bilingual domain. In contrast to the IA model, the BIA model (shown in Figure 4) can deal with words from two languages and therefore simulate visual word recognition in a bilingual. It implements bottom-up word activation in a language non-selective fashion (letters activate words from all languages). Language nodes were added to the original IA model structure that serve as a linguistic representation for language membership (indicating to which language a word belongs). A language node receives activation from all orthographic word representations belonging to that language. The language node, in turn, inhibits the orthographic representations of the words 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. Furthermore, because the basic structure of the BIA model

is no different from the IA model, it suffers from the same restrictions in word length due to position-specific coding of the letters.

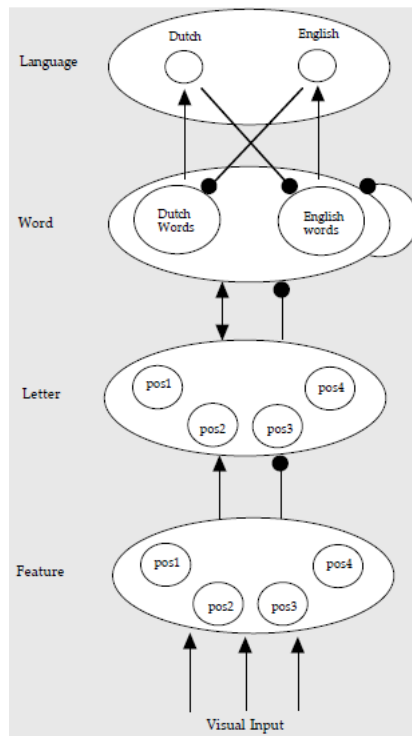


Figure 4. The Bilingual Interactive Activation model

#### *The BIA+ model*

The restrictions imposed on the BIA model as well as new empirical evidence on bilingual word recognition led Dijkstra and van Heuven (2002a) to develop an extension to the BIA model, referred to as the BIA+ model (Dijkstra & van Heuven, 2002a) (Dijkstra & van Heuven, 2002b). This model is displayed in Figure 5.

The earlier BIA model was extended with lexically phonological and semantic representations of the words. Furthermore, a task/decision system was introduced to allow the model to execute different tasks and to distinguish (non-linguistic) task-related effects from linguistic effects. Linguistic information from input signal or (sentence) context may affect the word identification system, while non-linguistic context information (e.g., participants'

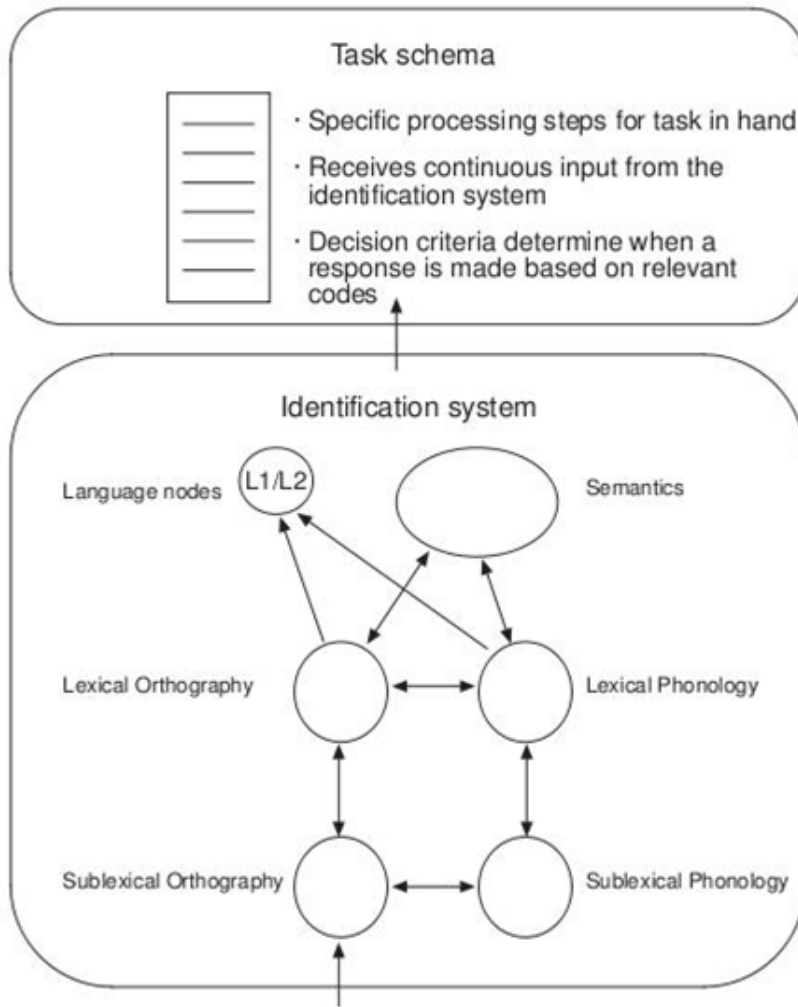


Figure 5. The BIA+ model

expectations and strategies) influences parameter settings in the task/decision system. 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. Because implementing lexical phonological and semantic representations computationally is quite difficult, for now the authors only provide a verbal analysis for these components of the model. As we will see later, the task/decision system proposed in the BIA+ model is similar to that used in the Multilink model developed in the present study.

### *The Revised Hierarchical Model*

According to the Revised Hierarchical Model (RHM) by Kroll and Stewart (1994) depicted in Figure 6, the two lexicons (of L1 and L2) of a bilingual are bi-directionally connected via lexical links. The lexical link from L2 to L1 is stronger than the link from L1 to L2, to reflect the way L2 was learned. During the acquisition of L2, bilinguals learn to associate every word form from L2 with its equivalent in L1 (e.g. ‘bike’ to ‘fiets’). This process forms a strong lexical-level association (Kroll & Stewart, 1994). The weaker connection from L1 to L2 reflects a lack of translation practice and the stronger lexical links from L2 to L1 reflect the bilingual’s ease of translation.

The lexicons are connected to their semantic representations (or concepts) via conceptual links. The connection from L1 to the concepts (represented by a solid line) is stronger than the connection from L2 (represented by a broken line) to the concepts. This reflects the larger familiarity of bilinguals with L1 word meanings, as L1 is their native language (Kroll & Stewart, 1994).

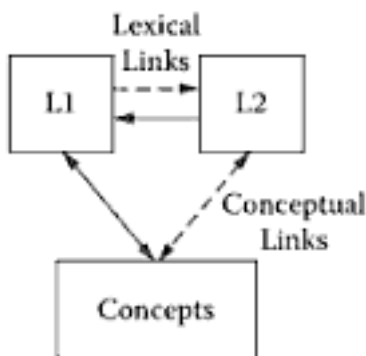


Figure 6. The Revised Hierarchical Model

According to this framework, there are two basic ways in which a bilingual can translate a word. The first option is that the bilingual perceives the word in one language, looks up its meaning, and then generates the associated word form in the other language. This mechanism has been called Concept Mediation (Potter, So, von Eckardt, & Feldman, 1984).

The second option is that the word form representation in one language ('bike') directly activates the representation of the translation equivalent in the other language ('fiets'). This mechanism has been referred to in the literature as Word Association (Kroll & Stewart, 1994).

RHM predicts an asymmetry in translation time. Translating from L1 to L2 should take more time than translation from L2 to L1, because the L1 to L2 translation is achieved by means of concept mediation, whereas L2 to L1 translation is achieved via word-association. Some translation studies, but not all, have confirmed this prediction.

### *WEAVER++*

WEAVER++ (Word Encoding by Activation and VERification) is a computational model designed to explain how humans plan and attentionally control the production of spoken words (Roelofs, 1997). It has been applied predominantly to monolingual data, but has more recently also ventured into the bilingual domain (see, e.g., (Roelofs, 2003) (Verhoef, Roelofs, & Chwilla, 2009)). The model is a "hybrid" model of human performance in the sense that it combines a declarative associative network and procedural rule system with spreading activation and activation-based rule triggering. Figure 7 shows an example of the declarative associative network. We will not go into detail about the WEAVER++ model. Note that in order to compute a mean response time distribution the model proposed in this study uses a hazard rate function and Luce choice rule similar to those proposed by Roelofs (1992) and used in WEAVER++. This procedure is described in more detail in the following sections.

### *Research questions*

Already at a relatively early stage of model development, one must decide whether the model being developed is, in principle, capable of simulating the processes for which the model is being devised. Such an investigation into the model's general adequacy might reassure the modeler that he or she is working in the right direction and may prevent wasting valuable time and resources. Because the new model should provide an account of

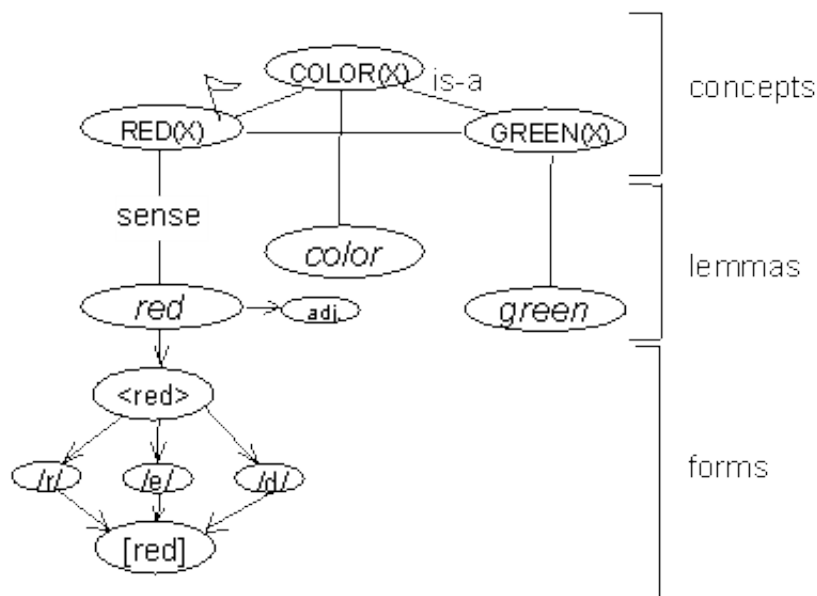


Figure 7. An example of the labeled associative network used by WEAVER++ to store ‘facts’ about words.

the complete translation process in line with general linguistic effects known in the field (e.g., effects of word frequency or word length), it needs to be able to show these linguistic effects in the first place. The present study provides such a preliminary investigation. Because the model in question is still at an early stage of development, we will inspect these effects at face value. In other words, we will simply investigate whether the proposed model is capable of showing general patterns of linguistic effects. After inspecting some basic effects, we will proceed to a more detailed comparison of two model-generated datasets to empirical data that resulted from recent studies. In these comparisons, we investigate whether the effects of orthographic similarity in the model-generated data are similar to those in the empirical data. More specifically, we will consider the direction of these effects. We will describe all investigated effects in more detail in the corresponding sections. To summarize, the questions we are attempting to answer in this study are presented in Table 1.



Question	Section
Is the model capable of showing	
Word frequency effects	Simulation study 1
Word length effects	Simulation study 2
Semantic priming	Simulation study 3
How does the model compare to empirical data:	
From a Lexical decision task	Empirical Data
From a Language decision task	Empirical Data

Table 1: Research questions addressed by this study

*Expectations*

Because we wish to apply a nested modeling approach, the model proposed in the present study is quite similar to the IA and BIA models (as will become evident in the following section). Because these earlier models have already shown to be capable of exhibiting the Word frequency effect, we expect the new model to show this effect as well. Innovative for the new model is its ability to process words of different length and semantics. This enables us to investigate Word length effects and Semantic priming effects. Although these abilities cannot be directly attributed to model components derived from the models discussed earlier, the present model was designed especially to exhibit such properties, so the presence of these effects is anticipated.

### Multilink: a new model for multilingual processing

In the present study, we propose a new computational model for multilingual processing called the Multilink model. Simulating the word translation process forms one special function of the model, which is capable of performing several other tasks as well (e.g. cognate recognition and semantic priming). In the following section, we will first present a global description of the model's structure and main processing components. Next, we will discuss these components and the currently implemented subcomponents in more detail. Because the model is in early development stages, not all subcomponents have been implemented yet, so in discussing them we will restrict ourselves to the implemented subcomponents. Finally, we will consider how input letter strings are processed by the model.

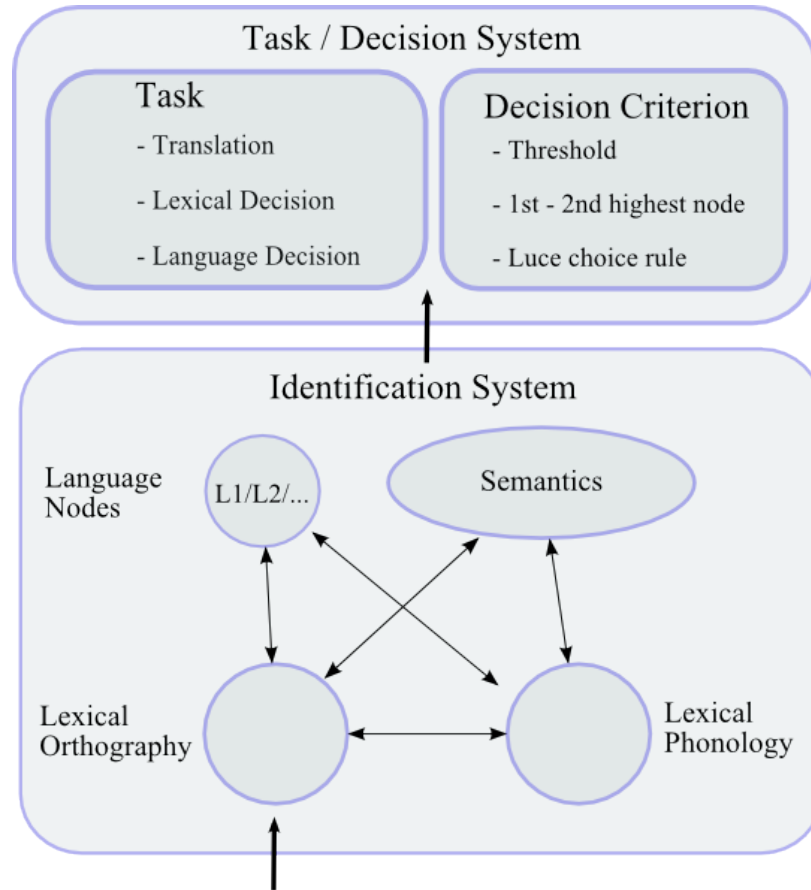


Figure 8. The Multilink model

*General description*

The Multilink model we are proposing in the present study is similar to the IA and BIA models discussed earlier in that it is an implemented computational model containing a localist connectionist network. The global structure of the model is actually more similar to the previously discussed BIA+ and RHM models. Like BIA+, the model is comprised of a task/decision system on top of an identification system, as can be seen in Figure 8. Later, we will discuss these components in more detail. Like in BIA+, Multilink’s distinction between an identification system and a task/decision system is in line with the view that performing a task (such as lexical decision) consists of two processing levels; a preconscious, automatic level followed by an attention-sensitive level. During the latter percepts are selected with reference to contextual factors and linked to a response relevant to the task.

The main difference with the BIA+ model is that Multilink has no sublexical (orthographic and phonological) layers. This frees the model from an important restriction that both the BIA(+) models as well as the original IA model by McClelland and Rumelhart (1981) suffered from, namely position specific encoding. The reader may recall that position specific encoding of the letters made it impossible for a model to process words of different lengths. Removing the sublexical layers frees us from position specific encoding and therefore endows us with the ability to process words of different lengths. A downside to this is that Multilink is unable to account for any sublexical effects such as the word superiority effect in letter perception (see e.g. (Harley, 2001)). Another consequence is that an alternative metric for linking an input letter string to stored lexical representations had to be devised. Note that in the earlier models, the sublexical layers would spread their activation to the lexical layers. This new strategy will be discussed in more detail in the following sections.

*Structural description*

In this section we will take a closer look at the structure of the components and subcomponents of the Multilink model. We will start by describing the Word Identification

System and then the Task/Decision System. For both components we will also consider the parameters involved.

### Word Identification System

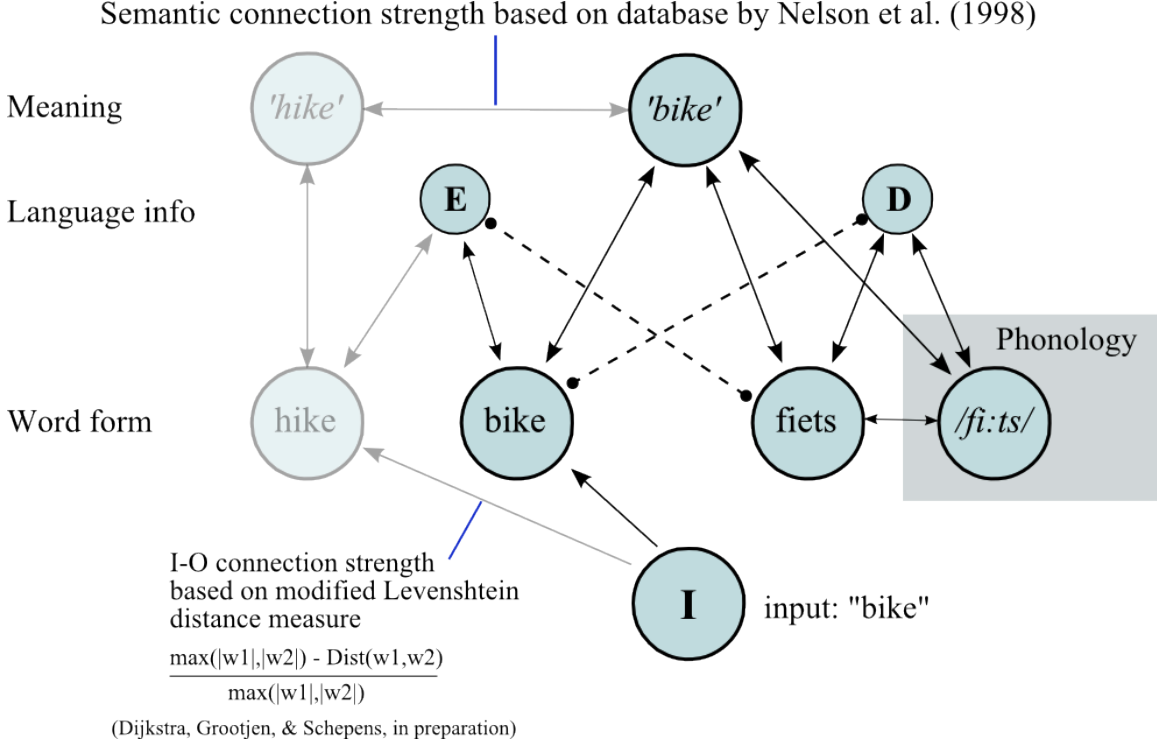


Figure 9. The Word Identification System of the Multilink model

The Word Identification System is a localist connectionist network containing several layers of symbolic representations, like in the IA and BIA models discussed earlier. The symbolic representations themselves form the nodes in the network, as can be seen in Figure 9. These nodes may receive and spread activation from and to other nodes as well as be inhibited by or inhibit other nodes. The activation of a node at a certain timestep is a function of the combined activation and inhibition it receives from other nodes (the net input). That function is referred to as the activation function of a node. The activation function used in Multilink is essentially the same as the one used in the IA model (McClelland & Rumelhart, 1981). Equation 0.1 shows how the net input at a given

timestep  $n_i(t)$  of a node is computed, where  $\alpha_{ij}$  is the weight of an incoming excitatory connection between node  $i$  and neighboring excitatory node  $j$ ,  $e_j$  is the activation of node  $j$ ,  $\gamma_{ik}$  is the weight of an incoming inhibitory connection between nodes  $i$  and  $k$  and  $i_k$  is the activation of node  $k$ . Equations 0.2a and 0.2b show how the net input contributes to the effect  $\epsilon_i(t)$  on a node, when the net input is positive and negative, respectively. In 0.2a  $M$  is the maximum activation of a node while in 0.2b  $m$  is the minimal activation of a node. In both cases  $a_i(t)$  is the current activation of the node. Finally, Equation 0.3 shows the activation function, in which  $\Theta_i$  is the decay rate of the node (pushing its activation back to the resting level) and  $r_i$  is the resting level activation of the node. The spreading of activation in the Word Identification System allows the Task/Decision System to select a response, based on the activations of individual nodes or node groups.

$$n_i(t) = \sum_j \alpha_{ij} e_j(t) - \sum_k \gamma_{ik} i_k(t) \quad (0.1)$$

$$\epsilon_i(t) = n_i(t)(M - a_i(t)) \quad \text{if } n_i(t) > 0 \quad (0.2a)$$

$$\epsilon_i(t) = n_i(t)(a_i(t) - m) \quad \text{if } n_i(t) \leq 0 \quad (0.2b)$$

$$a_i(t + \Delta t) = a_i(t) - \Theta_i(a_i(t) - r_i) + \epsilon_i(t) \quad (0.3)$$

On the implementational level, the model is constructed in a way that resembles the model's abstract components discussed earlier. First the identification system is constructed using a list of concepts. On each line of this list, for every language in the model, a word representing the concept is listed. This allows the model to create the semantic (conceptual) nodes and connect them to the appropriate orthographic nodes as shown in Figure 9. This concept list also allows for transcribed phonological representations of a concept. For each language in the list, a language node is created. A language node is excitatorily connected to

English	Dutch
COLOUR,85	KLEUR,97
SALT,44	ZOUT,41
THUMB,24	DUIM,25
DOCTOR,136	DOKTER,130
TOMATO,7	TOMAAT,2

Table 2: An example of the concept list used by Multilink to construct the Word Identification System

the orthographic word-forms belonging to the language represented by that language node, and inhibitorily connected to the word-forms belonging to the other language nodes. If, for example, the concept list contains orthographic word forms for the concept of ‘tomato’ in English and Dutch, five nodes are created; the concept node, an orthographic node for each language (TOMATO and TOMAAT) and finally the two language nodes (English and Dutch). For every additional concept the language nodes are maintained and only the concept and orthographic nodes are newly created.

Along with the word, the frequency of usage of said word in its corresponding language is specified. This can be any number, occurrences per million or subjective ratings for example, as long as the numbers are on the same scale for both languages. Table 2 shows an example of such a list. This particular example uses occurrences per million. For each of the languages the researcher can specify a number that is to be multiplied with the frequency rating for that language. This allows the researcher to adapt the frequency ratings to reflect a beginning or proficient multilingual. The underlying idea is that beginning bilinguals, for instance, would encounter and use words in their native tongue relatively more frequently than words in their second language. After being multiplied by this pre-determined value, the new frequency of usage ratings are used to create a ranking-list in which all orthographic representations of the words (of all languages) are sorted on frequency. From this ranking-list, the resting level activations of the orthographic nodes are determined. This gives more frequent words a head start in the activation spreading. The resting level activation is computed as presented in Listing 1, where:

- MIN\_REST is the minimal resting level activation parameter.
- RANK is the ranking of the word in the frequency-ordered list. (The higher the frequency the higher the RANK.)
- Math.abs() computes the absolute value.
- MAX\_REST is the maximum resting level activation parameter.
- MAX\_RANK is the highest rank found in the frequency-ordered list. (Rank 0 is the lowest rank.)

---


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


---

Listing 1 Computation of resting level activation.

When all concepts have been entered into the network, the semantic relations between concepts are looked up in the Free Association database created by Nelson, McEvoy, and Schreiber (1998). These values are multiplied with a pre-set parameter allowing the researcher to increase the relative strength of the semantic system. Note that semantic and associative links are actually not the same thing and that our model can, in principle, use any database which encodes relationships between concepts. We use a free association database because it has several advantages compared to other techniques and is considered a reliable method for measuring connection strength. See (Nelson et al., 1998) for an overview of advantages and shortcomings. We have used this particular free association database because to our knowledge it is the single largest database of free association ever collected. It was constructed from nearly three-quarters of a million responses by 6,000 participants to 5,019 stimulus words. Nelson et al. (1998) presented their participants with stimuli like BOOK -----, and asked them to write down, on the blank, the first word that came to mind that was meaningfully related or strongly associated to the presented word.

*Parameters.* Each node is characterized by three standard parameters that are constant and equal for all nodes.

- MIN\_ACT: the minimal activation of a node
- MAX\_ACT: the maximal activation of a node
- DECAY\_RATE: the decay rate of a node

Each node type (input, language, orthographic, phonological and semantic) has specific resting level activations:

- I\_rest
- L\_rest
- S\_rest
- P\_rest

For orthographic nodes, the resting level activation is computed from the frequency ranking. It therefore does not occur in the parameter list above. The value can, however, be kept within certain bounds by using the following parameters:

- MIN\_REST: minimal resting level activation
- MAX\_REST: maximal resting level activation

There are two types of connections in the identification system; excitatory and inhibitory connections. We call the weights of excitatory connections “alpha” weights and the inhibitory connections “gamma” weights. We use the first letter of the pool name to which nodes belong along with the alpha and gamma terms to refer to a specific type of connection. IO\_alpha, for example, refers to the excitatory connection from the input node to the orthographic nodes, while OO\_gamma refers to the inhibitory connection from the orthographic nodes to other orthographic nodes. Table 3 shows the types of connections in the present implementation of the Multilink model.

Connection types that are not computed automatically or disabled by default have weight parameters that can be set by the researcher. The parameters that are computed automatically can be adjusted in order to increase or decrease their effect by using the following multiplication parameters:

- IO\_multiplier



Weight	Comment
IO_alpha	Computed from modified Levenshtein distance.
<del>IO_gamma</del>	There are no inhibitory connections coming from the input node.
<del>OI_alpha</del>	There is no feedback to the input node.
<del>OI_gamma</del>	There is no feedback to the input node.
OS_alpha	
OS_gamma	Disabled by default.
SO_alpha	
SO_gamma	Disabled by default.
OL_alpha	
OL_gamma	
LO_alpha	
LO_gamma	
OO_alpha	Disabled by default.
OO_gamma	Disabled by default.
SS_alpha	Computed from the free association database.
<del>SS_gamma</del>	No inhibitory connections exist between semantic nodes.
LL_alpha	Disabled by default.
LL_gamma	Disabled by default.

Table 3: Connection weight parameters that exist in the Multilink model.

- SS\_multiplier

Furthermore, the “neighborhood” of a word can be adjusted by setting the minimal Levenshtein score parameter that we will discuss in section .

### *Task/Decision System*

In this section, we will take a closer look at the Task/Decision system, which consists of two main parts; the Task and the Decision Criterion. The task system specifies what the model output will be and to which component of the identification system the decision criterion will be applied. The decision criterion is used by the task system to decide which of the task-specific outputs will be produced. The model was designed in such a way that any task can be combined with any decision criterion. The decision criteria themselves can be applied to any group of nodes (the ONodes, LNodes, etc.).

The tasks currently implemented are language decision, lexical decision for the languages added to the model, generalized lexical decision and word translation. The decision

criteria currently available to the model are a threshold, the difference between the highest and the second highest node activations, and the Luce choice rule. It is relatively easy to add tasks and decision criteria. In the following subsections, we will give a short description of the tasks and decision criteria. We will start by describing the three tasks, then proceed to the three decision criteria, and finally, describe some parameters involved in the task/decision system.

*Lexical decision.* The model supports two types of lexical decision; ‘standard’ lexical decision and generalized lexical decision. In standard lexical decision the subjects are shown a word and asked whether the word belongs to a particular language or not. They are usually asked to press a YES or NO button as quickly as possible. In the case of the model, the output is either a YES or NO and the timestep on which this decision was made. The task can be applied to any of the languages loaded into the model. The model allows for two distinct ways of reaching the YES or NO decision. The first option is to put a decision criterion on the language nodes. The second option is to put the decision criterion on the orthographic nodes of the language of interest (Grainger & Jacobs, 1996).

Generalized lexical decision entails the same task, only the subject is not asked whether the word belongs to a particular language, but is asked whether it is a word at all. The same decision strategies are allowed as in normal lexical decision, so the decision criterion can be put on the language nodes or on the orthographic nodes. When applying the criterion on the orthographic nodes, the entire pool of orthographic nodes is used instead of only the pool belonging to the language of interest.

*Language decision.* In language decision subjects are shown a word and asked to which language the word belongs. For this task the model supports only one decision strategy; one of the decision criteria is applied to the language nodes. The output of the model is the name of the ‘winning’ language node and the timestep upon a decision was reached.

*Translation.* In this study, translation is considered reading or hearing a word in one language and pronouncing it in another. The present implementation of the model

does not fully support this task yet, due to the lack of phonological nodes. As a proof of concept, however, it is possible to present a word to the model and have it respond with the orthographic form in another language. Because the internal mechanism in the model is the same for orthographic forms as for phonological forms, it would be relatively trivial to implement this task in this model.

*Reaction time distributions.* As mentioned before, the Multilink model has the ability to compute mean reaction time distributions for each input instead of a single reaction time. The computation of the reaction time distribution is based on the retrieval latency computations proposed by Roelofs (1992). Figure 10 shows an example distribution for the word BIKE. A description of how Multilink computes these distributions can be found in Appendix C.

*Threshold.* The first of the decision criteria we will describe is the threshold criterion. This is a classic decision criterion also used in the IA and BIA models. The criterion is satisfied when the activation of a node in the pool reaches a predefined threshold value.

*1st - 2nd highest activations.* The second decision criterion used by Multilink is the difference between the node with the highest activation and the node with the second highest activation. The value of this difference is compared to a predefined threshold. If the value is greater than the threshold, the criterion is satisfied. The idea behind this criterion is that the ‘winning’ node should have a sufficiently higher activation than the runner-up in order to be recognized.

*Luce choice.* The final decision criterion carries a resemblance to Luce’s choice axiom (Luce, 1959). Mathematically, Luce’s choice axiom states that the probability of selecting item  $i$  from a pool of  $j$  items is given by  $P(i) = \frac{w_i}{\sum_j w_j}$  where  $w$  indicates the weight (a measure of some typically salient property) of a particular item. In Multilink, the result of the equation described above is not used directly as a selection probability, but is compared to some threshold. If it exceeds the threshold, the decision criterion is satisfied.

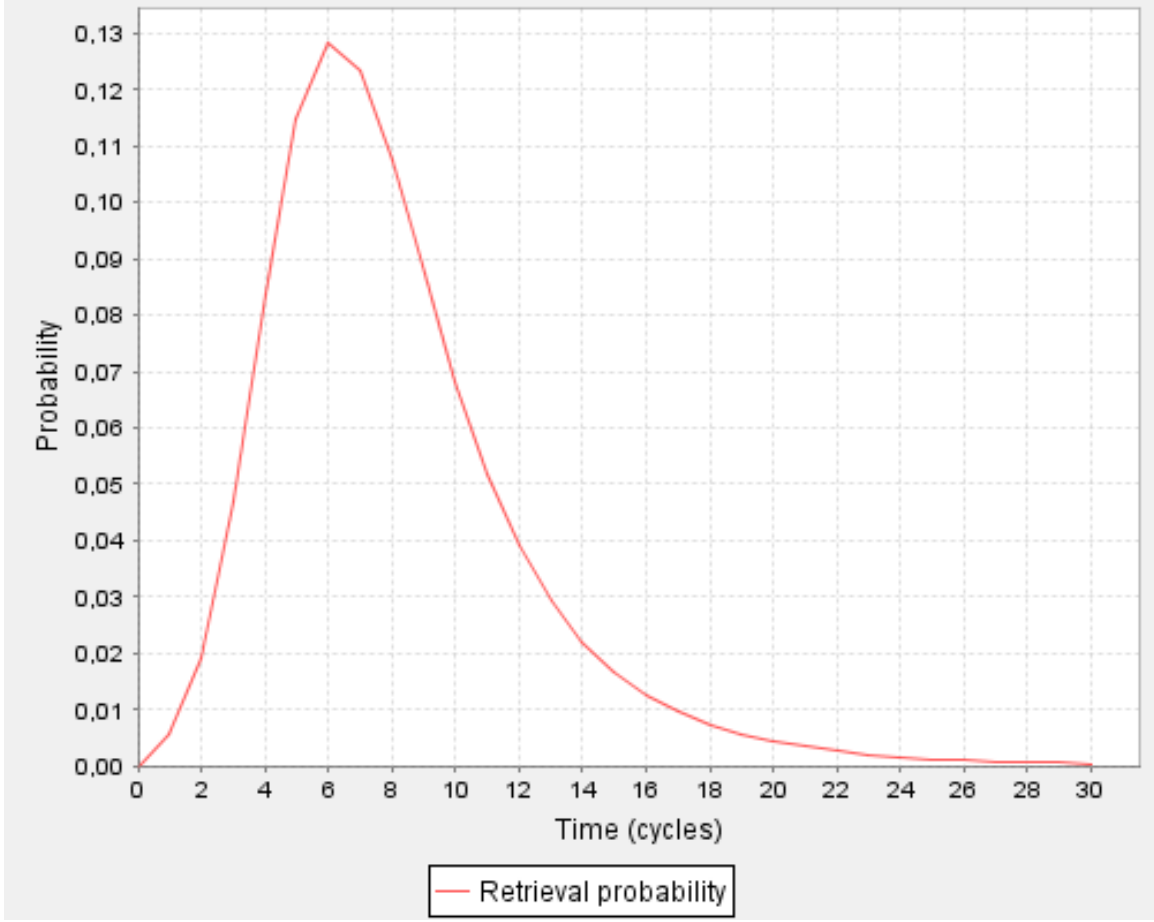


Figure 10. An example retrieval probability distribution for the word BIKE.

*Parameters.* We will now take a look at the parameters that are involved in the task/decision system. In fact, the task/decision system has only one real parameter, namely the criterion value. This parameter determines the value the recognition criterion should assume for a decision to be made. For instance, if the model is used with a generalized lexical decision task, a threshold decision criterion and a criterion value of 0.7, the model will output “YES” when an orthographic node reaches an activation threshold of 0.7. This parameter can also be used for the other tasks and criteria combinations, but it will have to be set specifically to match the task and criterion.

Aside from the criterion value parameter, there are two more parameters involved in

the task/decision system, namely the ‘timestep\_multiplier’ and the ‘timestep\_adder’. These parameters only affect the output’s unit of measurement, they do not influence the way the output is computed. They are added to and multiplied with the normal output of the model in order to allow the model to output reaction times in milliseconds instead of timesteps.

### *Processing description*

In this section, we will take a closer look at how input is processed by the Multilink model. First, we will discuss what makes up the input to the word identification system, then we will describe the word identification process, and finally, we consider in some details how translation works in this model.

#### *Input*

After the construction of the network, the identification system is ready to receive input. The input to the network consists of a single word. At each timestep, a different input may be provided to the system, allowing for the investigation of priming effects. A modified Levenshtein distance (Dijkstra, Grootjen, & Schepens, In preparation) is used to compute a similarity score between the input word and the orthographic representations in the model’s lexicon (see Equation 0.4). The Levenshtein distance was modified in order to normalize the measure on the length of the words. In the standard Levenshtein distance the length of the words is not taken into account, so two pairs of words with an equal amount 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, the pair ‘bike’ and ‘hike’ gets a score of 3/4 while the pair ‘finances’ and ‘financed’ gets a score of 7/8 even though they would have the same Levenshtein distance.

$$\frac{\max(|w_1|, |w_2|) - \text{Levenshtein}(w_1, w_2)}{\max(|w_1|, |w_2|)} \quad (0.4)$$

Based upon this score, the weights (strengths) of the connections between the input node and the orthographic nodes are computed. Listing 2 shows how these weights are

computed, where:

- score is the value of the modified Levenshtein distance.
- IO\_multiplier is a parameter setting which allows the researcher to adjust the network's input strengths.
- If the score is not greater than or equal to 0.5, the weight of the connection is 0.0

With regard to the spreading of activation in the network this initially results in a larger spread of activation to words similar to the input word than to words that are dissimilar. Upon every new input to the network, the connections from the input node to the orthographic nodes are recomputed.

---

```
if (score >= 0.5){IO_multiplier * score} else 0.0;
```

---

Listing 2 Computation of Input node to Orthographic node connection strength

Listing 2 shows that the strength of the Input to Orthographic nodes is set to 0.0 if the modified Levenshtein score is smaller than 0.5. We call this the minimal Levenshtein score parameter. We use this value because studies have shown that words that have less than 50% overlap do not influence each other in the identification process.

### *Word identification*

The word identification process consists of providing an input to the network and allowing it to spread activation from the input node to the other nodes in the identification system. How the activation spreads through the network depends greatly on the parameters of the model. An example will clarify the general procedure.

1. The (orthographic form of) the word “rat” is presented to the system.
2. The overlap between the input and the orthographic node (ONode) for “rat” (RAT) causes the activation of RAT to rise.
3. ONodes like CAT and HAT are also activated, but in lesser extent because of the smaller overlap.

4. In the Dutch/English multilingual version of the model, Dutch ONodes like KAT and RAT also become active.
5. The active ONodes activate their corresponding language nodes (LNodes) (Dutch and English).
6. The active ONodes activate their corresponding semantic nodes (SNodes) (the concepts rat, cat, hat, etc).
7. The active SNodes activate closely related concepts (according to the Free Association DB).
8. There is top-down feedback from the active SNodes and LNodes to their corresponding ONodes.
9. The activation spreads further until a certain decision criterion is met.

The identification system takes care of the identification process, but the decision which word must be recognized (on the basis of a decision criterion) is made in the Task/Decision system. The activation patterns of the nodes across time are used as input to the Task/Decision system.

### *Translation*

Translation can be accomplished by the model mainly because of the semantic nodes. As we have described in the example above, when an orthographic node is activated, it will automatically also activate its concept or meaning. When orthographic nodes from other languages share this meaning, they become active due to the top-down feedback from the semantic layer. If we would, for instance, give the English word ‘BIKE’ as input to the model, both the Dutch word ‘FIETS’ and its phonological form would also become active. Because there is attentional control over the translation process, the Task/Decision system is responsible for selecting the proper response. In this case, if the goal was to read ‘BIKE’, translate it from English to Dutch, and pronounce the translation, the Task/Decision system would select the Dutch phonological form as the response.

### Preliminary simulation studies

In this section, we will investigate whether some important effects known in the field of psycholinguistics can be reproduced by the Multilink model. Because the model is still in early development stages, our goal is to show that the model is capable of showing the investigated effects. In other words, we will be inspecting these effects more or less “at face value”.

#### *Simulation study 1: Word frequency*

How frequently a word is used is an important determinant of the speed and accuracy of word recognition. Words that are used on a frequent basis are recognized more easily and more quickly than less commonly used words. In other words, words with a higher frequency have lower reaction times. This effect has been found in a range of different tasks and is known to apply not only to visual word recognition, but also to auditory word recognition. The frequency effect does not just apply to words that have large differences in frequency, but also to words that have only slightly different frequencies. It is even argued that frequency is the single most important factor in determining response speed in the lexical decision task (see e.g. (Harley, 2001)).

#### *Goal*

As the word frequency effect is a robust and common effect, it is important that the Multilink model is able to account for this effect. Our goal is to see whether the effect does indeed occur in the simulation results of an English lexical decision task of the Multilink model. To limit this preliminary simulation study, we will only use the threshold criterion.

#### *Method*

We have run the model on a lexical decision task. The input consisted of 290 English words in a range of different frequencies. The written word frequency values were obtained from the work by Kucera and Francis (1967). Output was a response (YES or NO) and reaction time per input word. We offered only actual words to the model and it was



unnecessary to filter incorrect responses from the results as no errors were made by the model (no words were classified as non-words). We applied a univariate GLM analysis of variance on the output.

### Results

Correlations			
		RTModel	Freq
Pearson Correlation	RTModel	1.000	-.628
	Freq	-.628	1.000
Sig. (1-tailed)	RTModel		.000
	Freq	.000	
N	RTModel	290	290
	Freq	290	290

Table 4: Simulation study 1. Correlations between the reaction times predicted by the model (RTModel) and the written word frequency (Freq).

### Tests of Between-Subjects Effects

Dependent Variable: RTModel						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	
Corrected Model	48.067 <sup>a</sup>	1	48.067	187.157	Sig.	
Intercept	27762.137	1	27762.137	108095.634	.000	
Freq	48.067	1	48.067	187.157	.000	
Error	73.967	288	.257			
Total	42969.243	290				
Corrected Total	122.034	289				

a. R Squared = .394 (Adjusted R Squared = .392)

Table 5: Simulation study 1. Results of the analysis of variance with the reaction times predicted by the model (RTModel) as the dependent variable and the word frequency (Freq) as predictor.

Table 4 and Table 5 show the results of the statistical analysis of simulation study 1. RTModel is the variable containing reaction times predicted by the Multilink model. The Correlations table shows a significant correlation ( $p = .000$ ) of  $-.628$  between word frequency (Freq) and the reaction times predicted by the model. Table 5 shows a highly significant effect of word frequency on reaction time ( $p = .000$ ). The  $R^2$  value of  $.394$  indicates a strong effect.

### *Discussion*

The results show a significant and strong effect of word frequency on reaction time. The negative correlation reveals that higher frequency words have lower reaction times, which is in line with the common finding discussed above.

### *Simulation study 2: Word length*

The word length effect in visual word recognition suggests that longer words take longer to recognize than shorter words. This effect is not without controversy. One of the reasons the word length effects remain illusive is that there are three different ways to measure word length: the number of letters in a word, how many syllables, and how long it takes to say the word (Harley, 2001). Several studies have found an inhibitory effect of word length on reaction times in a lexical decision task using the number of letters as measure for the word length (Whaley, 1978) (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004), while others found no significant effect (Richardson, 1967) (Frederiksen & Kroll, 1967).

### *Goal*

Our goal is to show that the present model can be used for investigating word length effects, setting it apart from the IA and BIA models as they are not capable of investigating word length effects due to the position-specific encoding discussed earlier, and to see whether such an effect occurs in a lexical decision task with the default parameter settings shown in Appendix A.

### *Method*

We have run the model on a lexical decision task using the threshold criterion. Input was a random list of English words with lengths of four, five and six letters ( $N = 105, 93, 85$  respectively). Output was a response (YES or NO) and reaction time per input word. We have applied a univariate GLM analysis of variance on the output.

## Descriptive Statistics

Dependent Variable: RTModel			
Number of letters	Mean	Std. Deviation	N
4	12.0376121	.644513644	105
5	12.0552340	.620046357	93
6	12.1465216	.641562183	85
Total	12.1465216	.651885533	283

Table 6: Simulation study 2. Descriptive statistics of the reaction times predicted by the model (RTModel) and the number of letters in the input words.

## Tests of Between-Subjects Effects

Dependent Variable: RTModel							
Source		Type III SS	df	Mean Square	F	Sig.	Partial Eta <sup>2</sup>
Intercept	Hypothesis	41519.564	1	41519.564	12450.475	.000	1.000
	Error	6.676	2.002	3.335 <sup>a</sup>			
nLetters	Hypothesis	6.691	2	3.346	8.279	.000	.056
	Error	113.146	280	.404 <sup>b</sup>			

a. .996 MS(nLetters) + .004 MS(Error)

b. MS(Error)

Table 7: Simulation study 2. The results of an analysis of variance with the number of letters in the input words (nLetters) as predictor for the reaction times predicted by the model (RTModel).

*Results*

Table 7 shows a significant effect of the number of letters in a word on the reaction time as predicted by the Multilink model ( $p < .000$ ). The Partial Eta Squared of .056 suggests a weak effect. Figure 11 shows a plot of the estimated marginal means.

*Discussion*

The simulations show that the Multilink model can be used to investigate word length effects. Although the simulations show a significant but weak effect of number of letters on reaction time predicted by the model, we cannot completely exclude that an intermediate variable might be involved, such as the number and frequency of neighbors the different classes of words have (longer words might have more high frequent neighbors). In future

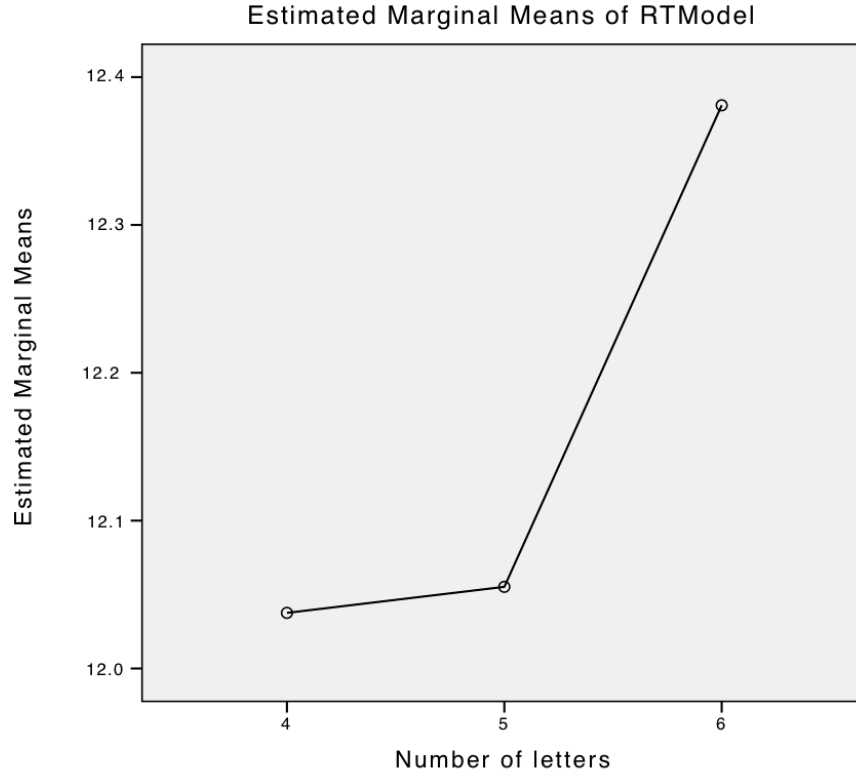


Figure 11. Simulation study 2 results: estimated marginal means

research, we suggest to control the input list for frequency and neighborhood size in order to investigate to what extent this effect is directly caused by word length.

### *Simulation study 3: Semantic priming*

Meyer and Schvaneveldt (1971) showed that the identification of a word is made easier if it is immediately preceded by a word related in meaning. For example, we are faster to decide that ‘doctor’ is a word if it is preceded by the word ‘nurse’ than if it is preceded by a word unrelated in meaning, such as ‘butter’, or than if it is presented in isolation. This effect, commonly known as the semantic priming effect, is a robust and widely examined effect. The largest semantic priming effects are found in lexical decision (Neely, 1991). Semantic priming can only be demonstrated and investigated by models that have some kind of ‘knowledge’ about the semantics of words and their relations. This is a property

the IA and BIA models lack. The Multilink model, however, does possess this knowledge by integrating the free association database (Nelson et al., 1998), into the identification system. This allows us to demonstrate semantic priming effects in model simulations.

### *Goal*

The goal of this simulation study is not to provide quantitative results on semantic priming effects in the Multilink model, but rather to give a proof of existence by showing an example of semantic priming effects simulated by the model.

### *Method*

Again, we used the threshold decision criterion to simulate lexical decision. In this study we will illustrate the semantic priming effect in lexical decision by means of a toy problem consisting of only three words: doctor, nurse, and purse. In the current model configuration, only nurse and doctor are semantically related, while nurse and purse and doctor and purse are not. First, we will only provide the target ‘DOCTOR’ as input to the model and then we will prime the target with both ‘NURSE’ and ‘PURSE’ and compare the results.

### *Results*

The graph in Figure 12 shows the model output for the unprimed input of the word ‘DOCTOR’. The word is recognized as being a word on interpolated timestep 12.28. We can see from the activations of the words nurse and purse in the same graph that nurse receives some top-down feedback from the semantic layer as the semantically related concept of nurse is activated by that of doctor.

The second figure (Figure 13) shows the model output for the word ‘DOCTOR’ primed by ‘NURSE’. Because the actual input of ‘DOCTOR’ now occurs 3 timesteps later than in Figure 12, we subtract 3 timesteps from the reaction time. As we can see, ‘DOCTOR’ is recognized more quickly than in Figure 12 (11.90 versus 12.28). We can also see that, because of the orthographic similarity between purse and nurse, ‘PURSE’ also receives

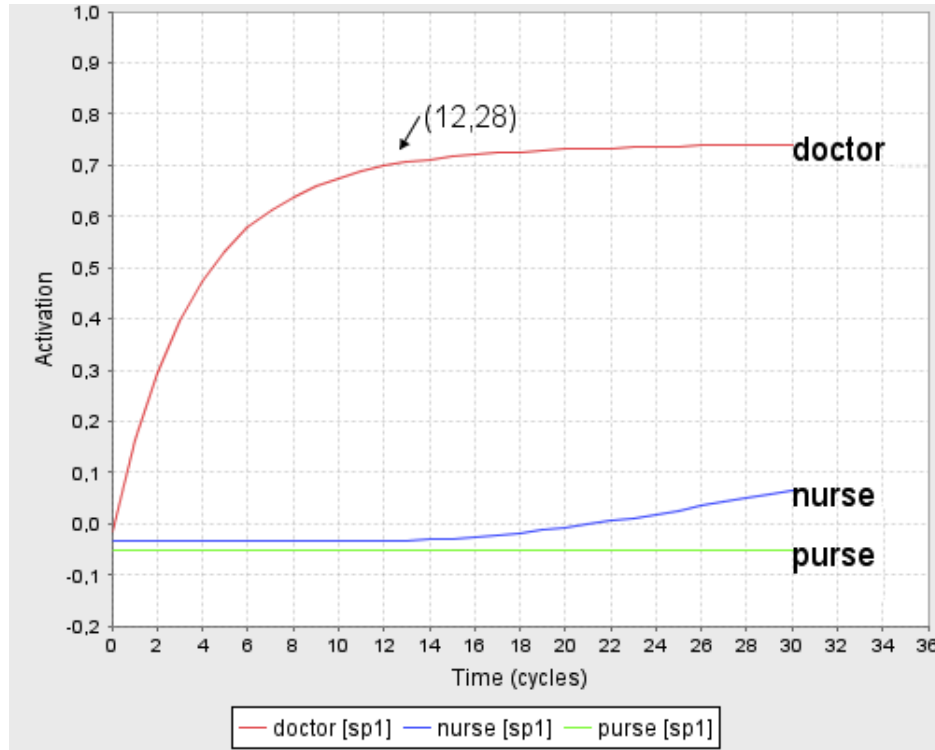


Figure 12. Simulation study 3: the (unprimed) word “DOCTOR” is recognized on timestep 12.28

activation from the input of ‘NURSE’, but then starts decaying back to its resting level activation.

Figure 14 shows us the recognition of ‘DOCTOR’ when primed with the semantically unrelated word ‘PURSE’. In this case, the reaction is still slightly faster than in the first case (12.09 versus 12.28). This is due to the orthographic similarity between purse and nurse. The orthographic node for nurse receives activation from the input of ‘PURSE’ and activates its concept nurse. Nurse then activates the concept doctor, which feeds back its activation to the orthographic node ‘DOCTOR’, causing a slightly faster response. It is, however, slower than the response in the second case (Figure 13), because ‘NURSE’ does not spread as much activation to ‘DOCTOR’ as in the second case.

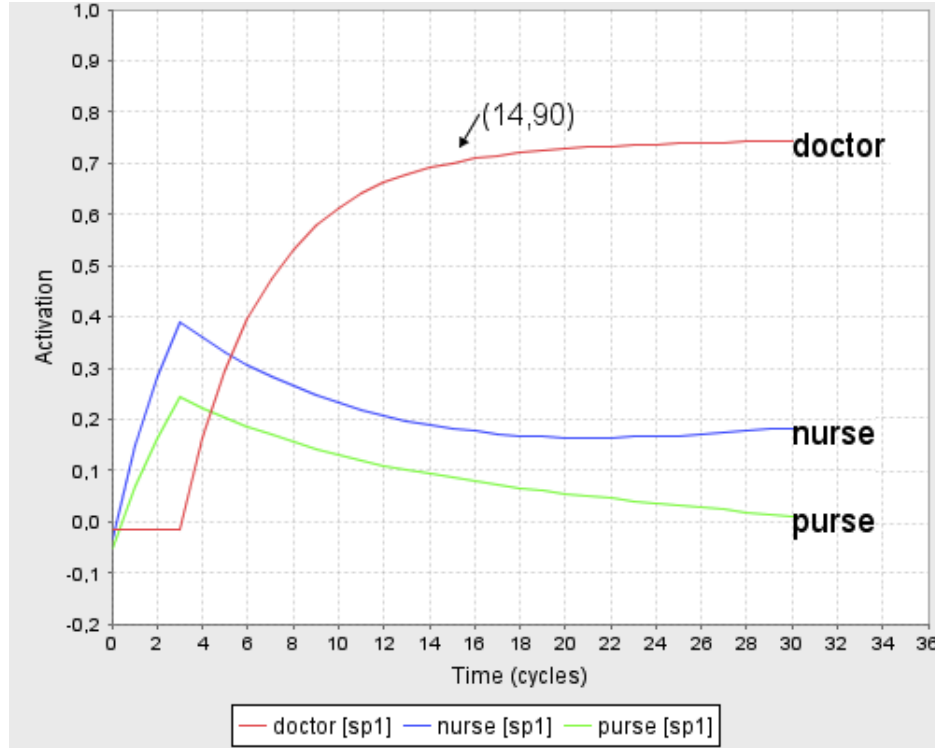


Figure 13. Simulation study 3: the word “DOCTOR” primed with “NURSE” is recognized on timestep  $14.90 - 3 = 11.90$

### Discussion

The results of this simulation show the model’s potential to elicit semantic priming and other semantically related effects, allowing us to gain insight in the underlying processes. The results in Figure 14, for instance, show immediately that the model may come up with new predictions that we could test using empirical data. One prediction in this case could be that priming a target with a semantically unrelated word that has a large similarity to a semantically related word may still lead to a facilitatory effect. It should be noted that the specific results gathered in this simulation are greatly dependent on the model parameters. The effects of these parameters should be investigated in further research.

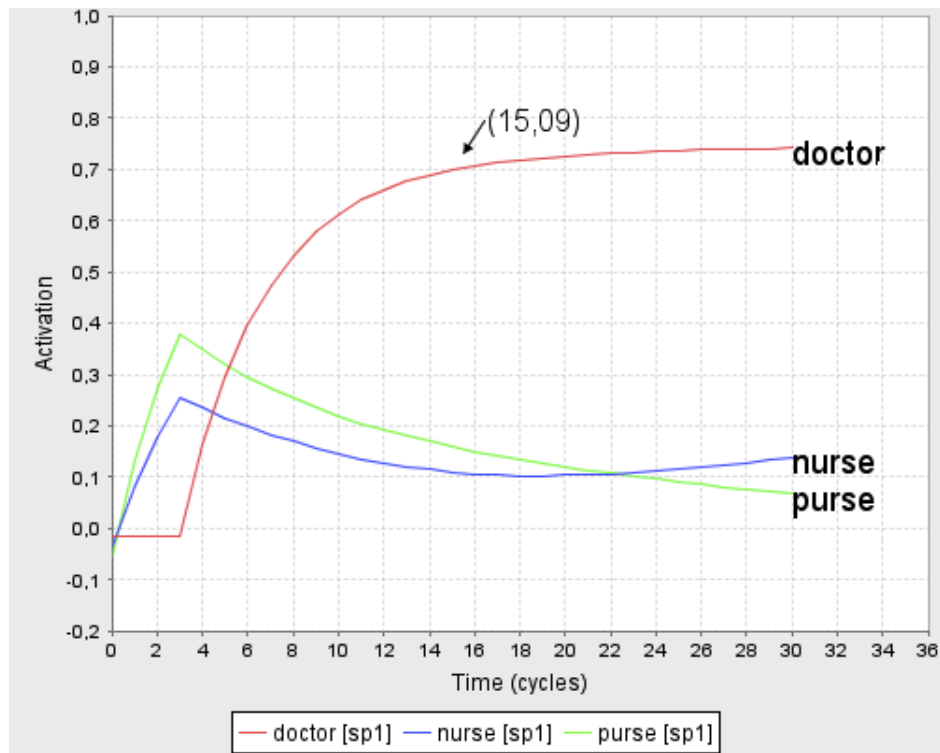


Figure 14. Simulation study 3: the word “DOCTOR” primed with “PURSE” is recognized on timestep  $15.09 - 3 = 12.09$



## Empirical Data Studies

### *Goal*

In the present study, M. Sappelli and myself compared simulation data produced by the Multilink model to language decision and English lexical decision data reported in Dijkstra, Miwa, Brummelhuis, Sappelli, and Baayen (2010). Furthermore, we examined the influence of different types of decision criteria on the fit of the model data with the empirical datasets.

### *Method*

We tested the performance of the model with respect to two tasks, namely, English lexical decision and language decision. In the lexical decision task, the output of the model is either a ‘YES’ or ‘NO’ response to a presented input word (or non-word) along with the timestep on which the model came to that decision. In the language decision task, the output of the model is the language membership of the input word and the timestep on which the model reached this decision. The decision is made by the model based on the selected decision criterion.

For each task, we ran the model three times, using the three different decision criteria available in the Multilink model, and compared the results to some of the results from Dijkstra et al. (2010) (shown in Figure 15 and Figure 16). The first criterion used, was the threshold criterion, in which the recognized word is the word that reaches the predefined threshold first. The second criterion used, was the criterion in which the highest activation is compared to the activation of the second highest, and when the difference between the two is more than a predefined value, the word is recognized (referred to as 1-2 difference). The final criterion was the Luce choice rule, which calculates the ratio between highest activation and the activation of the whole pool, and when this ratio is above a predefined value, the word is recognized.

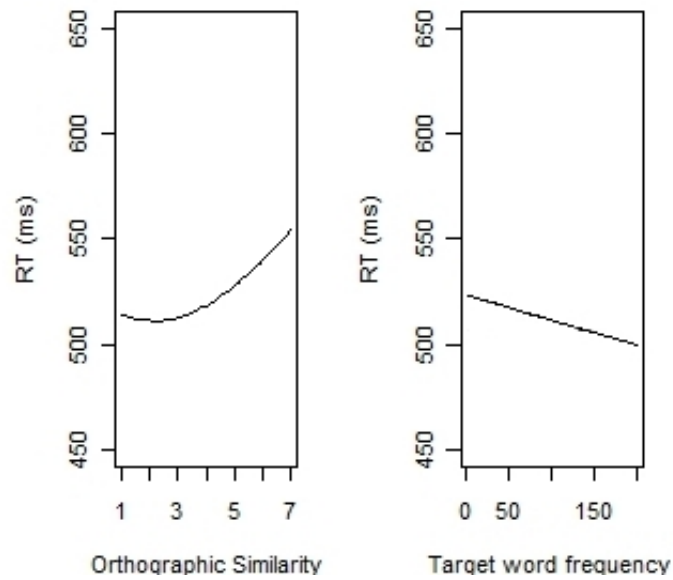


Figure 15. Results of the Language Decision experiment. It shows the shape of the effects of orthographic similarity and target word frequency on the reaction time in human participants.

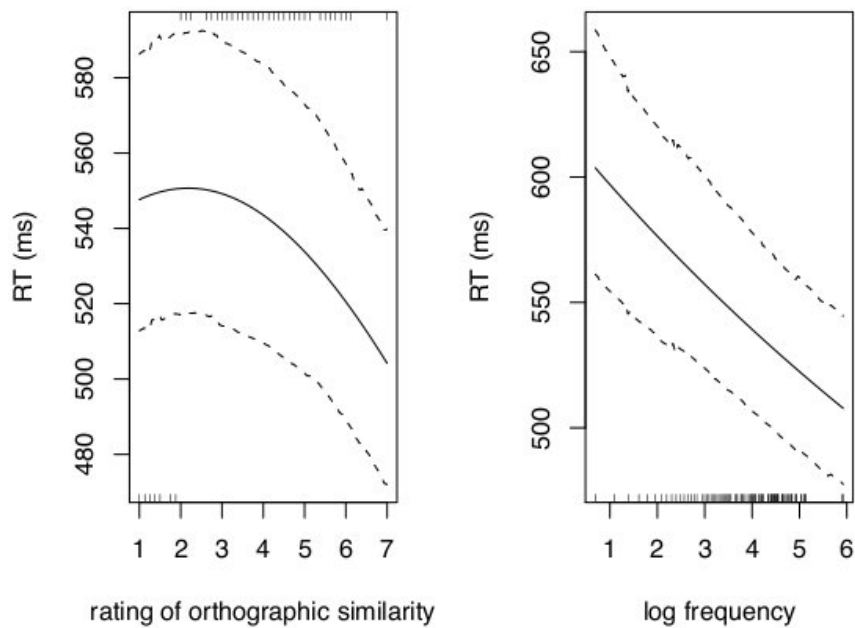


Figure 16. Results of the Lexical Decision experiment. It shows the shape of the effects of orthographic similarity and target word frequency on the reaction time in human participants.

### *Stimuli*

For the language decision task, the same stimuli were used as in the language decision experiment by Dijkstra et al. (2010). The list of input words is shown in Appendix B. For the english lexical decision task, no non-words were presented to the model and only the English words of the list used by Dijkstra et al. (2010) were used. The list of input words is shown in Appendix B.

### *Parameter settings*

The model mainly incorporated the parameters from the the IA and BIA model. The specific criteria values were either based on literature (the thresholds used in the IA and BIA models) or determined by means of trial and error. Because the lexical access in the Multilink model is different from the access in the (B)IA model, in the sense that there are no sublexical layers, there is a new IO-parameter that was manually determined. After trying out several parameter settings, the final model parameters did not include inhibition from orthographic nodes to opposite language nodes, but did include strong facilitation to language nodes and weak facilitation and normal inhibition from language nodes to orthographic nodes. An overview of the parameters used in this study is presented in Appendix A.

## *Results*

### *Language Decision*

Figure 17 shows the predictions the model made in the language decision task. Only the threshold criterion showed a significant non-linear inhibition effect of orthographic similarity on reaction time ( $p=0.03$ , non-linearity  $p=0.01$ ). This was comparable to the experimental data although the correlation was only 0.02 and the effect was not as strong as found in the experimental data. The frequency effect predicted by the threshold function was also significant and comparable to empirical data ( $p<0.01$ ). For the 1-2 difference, there was a remarkably high error rate of 57%. The 1-2 difference did not predict non-linear effects of

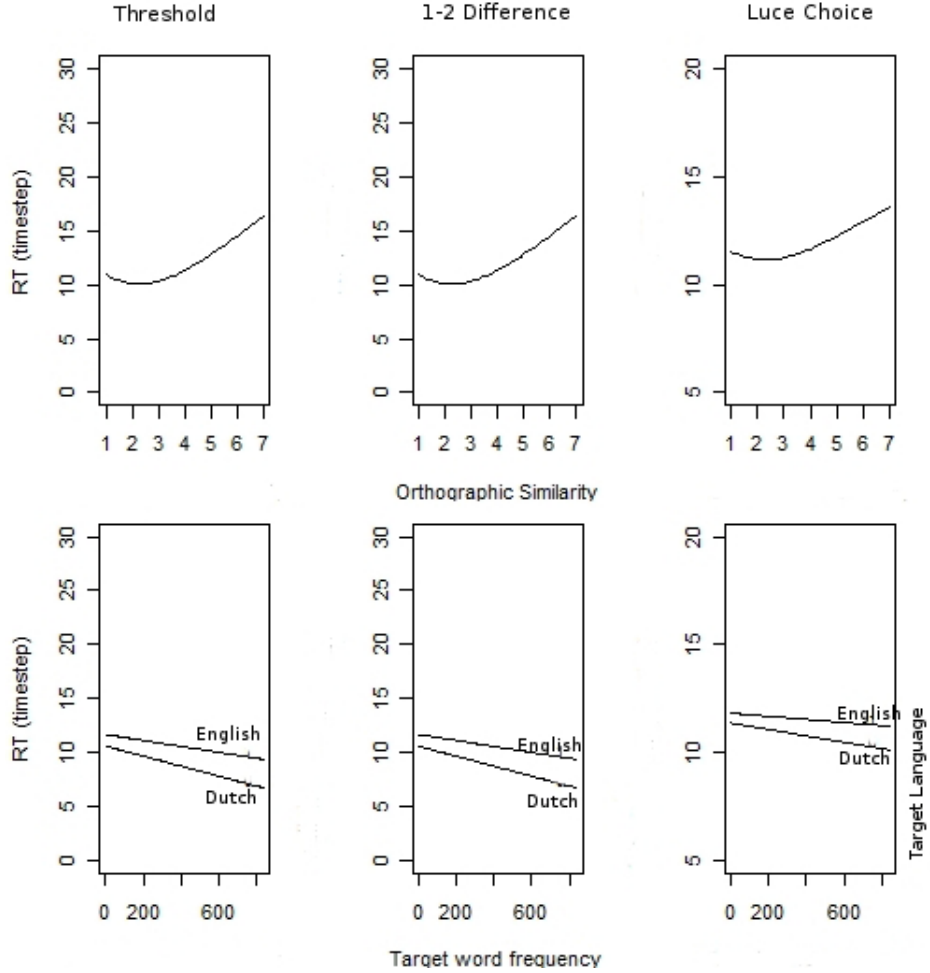


Figure 17. Simulation results of Language Decision without Orthography-Language inhibition

orthographic similarity, in fact, there was no significant effect of orthographic similarity at all ( $p=0.49$ ). The criterion did show a significant effect of word frequency comparable to that of the empirical data ( $p=0.03$ ). Finally, the Luce choice rule did not predict any effect of orthography, but did show a small but significant effect of frequency.

In an earlier test with different parameter settings, including inhibition from orthography to language, we found different results. In these simulations, all three decision criteria showed the same non-linear effect of orthographic similarity on reaction times as found in the experimental data. These effects were strongly significant ( $p<0.0001$ ). Also, the thresh-

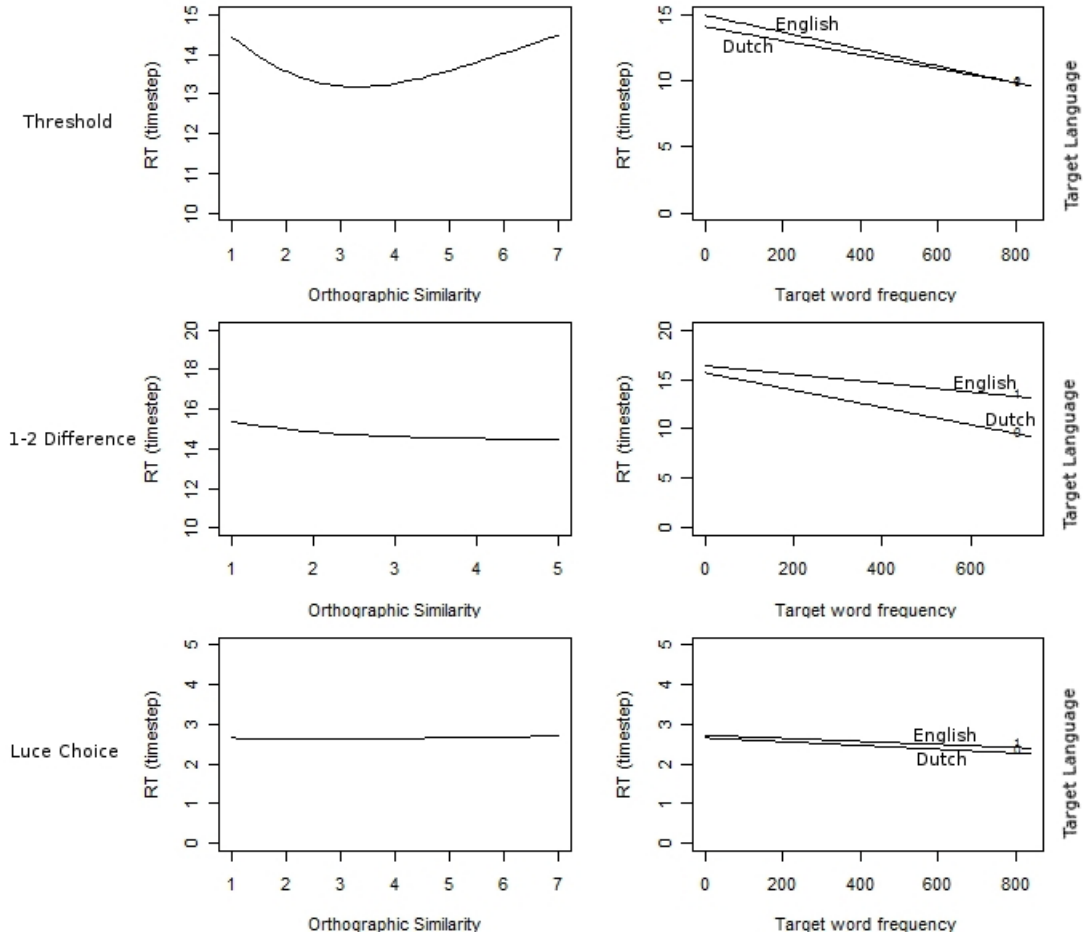


Figure 18. Simulation results of Language Decision with Orthography-Language inhibition

old criterion and 1-2 difference criterion yielded exactly the same results. This is because in this scenario, with the decision criteria applied to the language nodes, there are only two active nodes. Of these nodes, only one reaches an activation higher than 0.0, resulting in the same decision results as in the threshold criterion. These two criteria showed a significant overall frequency effect ( $p < 0.0001$ ), but no significant interaction effect, which is in correspondence with the experimental data. However, the shape of the effect seems to be different, as can be seen in Figure 18. Furthermore, no significant word frequency effect was found for the Luce choice criterion, although the shape of the effect seems to be the same as for the other criteria. The correlation of the threshold and 1-2 difference criteria with the experimental data is 0.197, based on 482 data points. This is much higher than

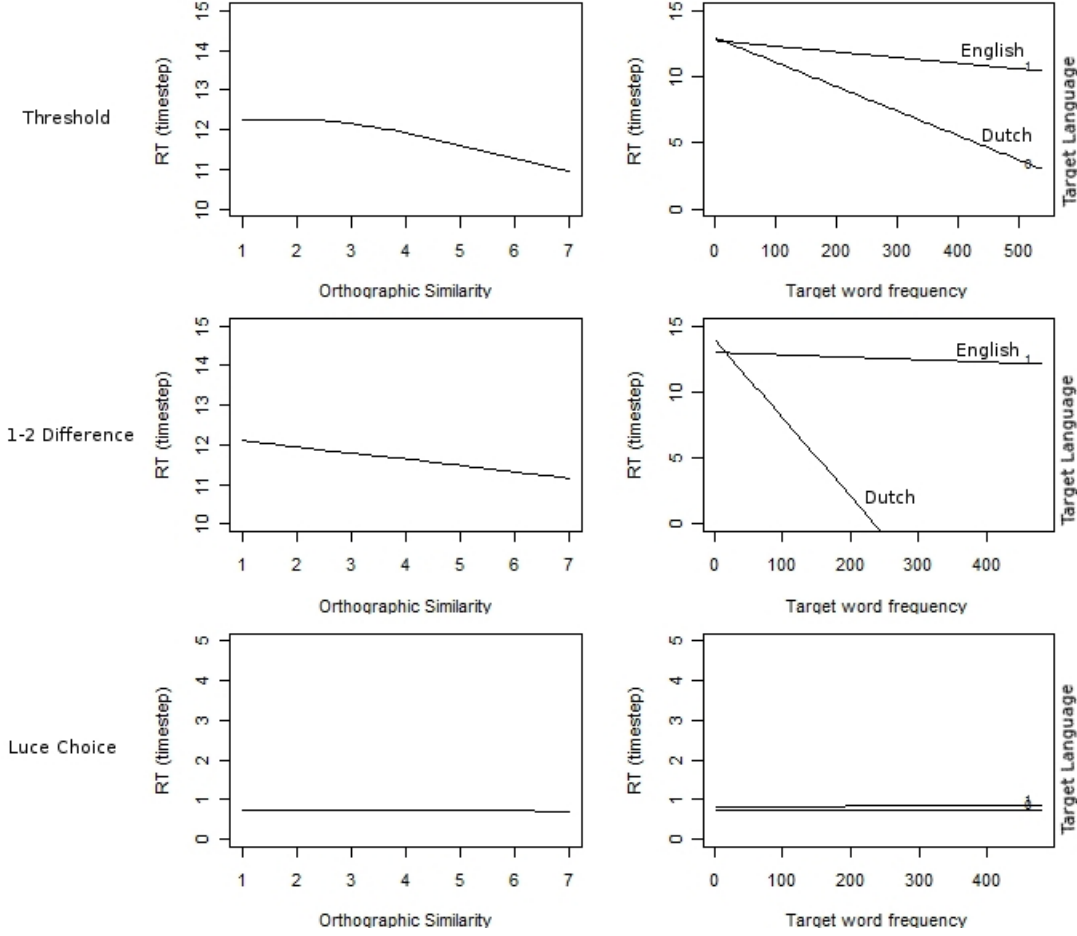


Figure 19. Simulation results of Lexical Decision without Orthography-Language inhibition

the correlation found without orthographic-language inhibition.

On the language decision data with orthography-language inhibition, using the threshold or 1-2 difference criterion, the model had an error-rate of 8.7% (46 errors). Of these errors 35% resulted in no response at all, because of too much competition between the choices and 50% of these no-response cases included identical cognates. Of the remainder, the majority of errors was made on cognates (66%) and most of the errors were made on English words (63%). For the Luce choice criterion the error-distribution is a little different. There was an error rate of 8.3% (44 errors) of which 20% yielded no response. Of these no-response cases, 90% of the inputs were identical cognates and the remaining 10% were nearly identical cognates (1 letter difference). Of the remaining errors 74% were cognates.

Using this criterion the model also predicted more mistakes on English than on Dutch words (63%).

Interestingly, the model without O-L inhibition showed fewer errors for the threshold criterion (6.4%), but shows a higher error-rate for the other criteria (respectively 57% for the 1-2 difference and 9% for the luce choice rule). The results show that enabling O-L inhibition yields a better overall performance.

### *Lexical Decision*

Using the model with O-L inhibition, we found none of the expected effects for the lexical decision task. We therefore continued the simulations using the model without O-L inhibition, although it provided slightly worse results for the language decision task. In the lexical decision task we found a strongly significant non-linear effect of orthography ( $p < 0.0001$ ) which reveals a facilitatory effect of similarity (Figure 19). Only the threshold function predicted this effect, although the 1-2 difference did predict a linear effect of orthographic similarity ( $p = 0.05$ ). The threshold criterion was the only criterion that predicted a significant effect of word frequency ( $p < 0.0001$ ), the effect predicted by the model is comparable to the effect Dijkstra et al. (2010) found. No significant interaction effect with target language was observed ( $p = 0.1$ ). Error-rates in the lexical decision task were very low for the threshold criterion (0%) while very high for 1-2 difference (73%) and Luce choice (47%). This suggests that only the threshold criterion is applicable for these tasks and parameter settings.

### *Discussion*

The Multilink model showed to be efficient in simulating the non-linear effect of orthographic similarity between Dutch and English words. There was a clear influence of type of decision criterion in the English lexical decision task, but no such influence was found for the language decision task, because for the latter the criteria were translatable into each other. The Luce choice rule showed not to be a good criterion for word recognition, because it led to artificial results and had a low significant predicting value. Furthermore, the 1-2

difference criterion was also not a good criterion because of its high error-rates. Cognate effects were found to be highly dependent on the task, implying that although the basic word recognition system may be the same in different situations, there is a large influence of the task at hand. The results obtained in this study are somewhat dependent on the model parameters. It is likely that a better fit of the empirical data can be obtained by adjusting the model parameters.



## General Discussion

In this paper we developed the Multilink model. In this section, we will first give a short summary of the structure and processing aspects of this model. Next, we will discuss the results of the preliminary simulation studies and empirical data studies. Then, we will discuss some of the advantages and disadvantages of the model, followed by some future research suggestions and a conclusion.

*Structure.* We introduced the two main components of the model, namely, the Word Identification System and the Task/Decision System. The Word Identification System is a localist connectionist network that is built using a list of concepts, containing orthographic and phonological forms of a word in one or more languages. Using this concept list, semantic, orthographic, and phonological nodes are created and connected to each other. For each language in the concept list a language node is created and connected to orthographic and phonological nodes belonging to that language.

The Task/Decision System is a rule based system that uses the activation patterns of nodes in the Word Identification System to make a response decision based on a given Task and Decision Criterion. Three main tasks are currently implemented, namely, lexical decision, language decision, and word translation. These tasks can be used in conjunction with three Decision Criteria, namely, a threshold, the difference between the first and second highest activations, and the Luce choice rule.

*Process.* When a stimulus word is presented to the Multilink model, activation spreads to the orthographic nodes. How fast the activation flows to a particular orthographic node depends on the similarity of the input word to the word represented by that node. From the orthographic nodes, activation spreads further across the network. Translation is allowed via concept mediation: activation flows between semantically related nodes and is fed back to orthographic and phonological nodes that are connected to that concept. The spreading of activation in the Word Identification System allows the Task/Decision System to come to a decision and select a response.

*Preliminary simulation studies.* Our preliminary simulation studies indicate that the model shows promising properties. It exhibits the robust word frequency effect and is capable of simulating word length effects and semantic priming effects. This ability sets the model apart from the models discussed earlier in this study, as they are generally not capable of processing words of different length or semantics. We have not yet investigated to what extent the effects exhibited by the model are similar to those in empirical data, so this remains to be examined in future research.

*Empirical data studies.* The empirical data studies have shown that the Multilink model is quite capable of modeling the observed non-linear effects of orthographic similarity. In the process we gained insight in the applicability of the decision criteria. We concluded that the threshold criterion was the best recognition criterion for the tasks and parameters applied in the empirical data studies.

*Advantages and disadvantages of the model.* The present study has shown how the absence of sub-lexical layers allow the model to process words of different lengths. This great benefit is also one of the model’s weaknesses. In the IA and BIA(+) models discussed earlier, the sub-lexical layers allow the models to exhibit all sorts of sub-lexical effects, such as the word superiority effect. It is impossible for the Multilink model to model such effects.

We have also seen how the use of the free-associations database by Nelson et al. (1998) allows the model to build a semantic layer and connect related concepts. This is a great advantage of the model, because it also allows translation to take place via concept mediation. Furthermore, it allows for the investigation of a whole range of interesting phenomena that are associated to bilingualism and multilingualism, such as cognate effects. However, using a language-specific free association database also has its drawbacks. One is that associations may not only depend on conceptual links, but also on form links (cf. the Dutch expression ‘huisje, boompje, beestje’, where the link between the words is not semantic but associative). In addition, there may be culture-specific associations, which should be addressed in future research.

*Other future research.* As we have discussed previously, the Multilink model currently has no implemented phonological layer. One of the primary objectives for future research is to extend the model with such a layer allowing it to process phonological word forms as well as orthographic ones. The inclusion of the phonological layer would bring the model closer to the final objective; the modeling of the complete translation process.

In sum, Multilink produces a wide range of promising results, even though it is as yet in an unrefined early development stage. A major step forward in improving the model’s predictions is to fit the model parameters to empirical data. In order to optimize the core model for different task situations, finding a ‘universal’ set of parameter settings would be desirable. Although it is not necessarily the case that such a universal parameter set exists, one could attempt to find one using automated parameter fitting techniques on all tasks the model is designed to model. One could, for instance, apply a multi-objective evolutionary algorithm to search for an optimal parameter set for lexical decision, language decision, and translation tasks.

*Conclusion.* We have discussed several options for future research that can still contribute to the improvement of the Multilink model. Considering the complexity of the word translation process, already in it’s early stages the Multilink model brings us a lot closer to modeling this complex human feat.

## References

- Balota, D., Cortese, M., Sergent-Marshall, S., Spieler, D., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, 133(2), 283-316.
- Dijkstra, A., Grootjen, F., & Schepens, J. (In preparation). *Cognate distributions across six european languages*.
- Dijkstra, A., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, In Press, *Corrected Proof*, -.
- Dijkstra, A., & van Heuven, W. (2002a). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5, 175-197.
- Dijkstra, A., & van Heuven, W. (2002b). Modeling bilingual word recognition: Past, present, and future. *Bilingualism: Language and Cognition*, 5, 219-224.
- Frederiksen, J., & Kroll, J. (1967). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology : Human Perception and Performance*, 2, 361-379.
- Grainger, J., & Jacobs, A. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, 103, 518-565.
- Grosjean, F. (1981). Sign and word recognition: A first comparison. *Sign Language Studies*, 32, 195-220.
- Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism*. Cambridge: MA: Harvard University Press.
- Harley, T. (2001). *The psychology of language*. Sussex: Psychology Press Ltd.
- Kroll, J., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149-174.
- Kucera, & Francis. (1967). *Computational analysis of present-day american english*. Providence: Brown University Press.
- Luce, R. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. an account of basic findings. *Psychological Review*, 88, 375-407.
- Meyer, D., & Schvaneveldt, R. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227-235.

- Neely, J. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), (p. 264-336). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Nelson, D., McEvoy, C., & Schreiber, T. (1998). *The university of south florida word association, rhyme, and word fragment norms*. <http://www.usf.edu/FreeAssociation/>.
- Potter, M., So, K., von Eckardt, B., & Feldman, L. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 23, 23-38.
- Richardson, J. (1967). The effects of stimulus attributes on latency of word recognition. *British Journal of Psychology*, 67, 315-325.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107-142.
- Roelofs, A. (1997). The weaver model of word-form encoding in speech production. *Cognition*, 64, 249-284.
- Roelofs, A. (2003). Shared phonological encoding processes and representations of language in bilingual speakers. *Language and Cognitive Processes*, 18, 175-204.
- Romaine, S. (1995). *Bilingualism*. Oxford: Basil Blackwell.
- van Heuven, W., Dijkstra, A., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, 39, 458-483.
- Verhoef, K., Roelofs, A., & Chwilla, D. (2009). Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition*, 110, 84-99.
- Whaley, C. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 17, 143-154.

## Appendix A

Default parameters

## Appendix B

List of words used in simulations

*Simulation study 1*

*Simulation study 2*

*Simulation study 3*

*Empirical data study 1*

*Empirical data study 1*

## Appendix C

Calculation of reaction time distributions