Radboud University Nijmegen

Bachelor Thesis

Artificial Intelligence and Linguistics

Similarity-dependent cognate inhibition effects in language decision

Author:
Maya Sappelli
s0513504

Supervisor:
Dr. Ton Dijkstra

August 4, 2009

Similarity-dependent cognate inhibition effects in language decision

Maya Sappelli

Radboud University

Nijmegen, The Netherlands

04-08-2009

Abstract

We examined how the cross-linguistic similarity of cognates affects bilingual word recognition in a second language. In a language decision task, Dutch-English bilinguals processed cognates with varying orthographic overlap ratings of their English and Dutch readings (e.g., *"night - nacht"* vs. *"tennis-tennis"*). Relative to non-cognates, a non-linear inhibition effect was found on the reaction times for cognates that increased from similar to identical cognates.

The results are interpreted as evidence for competing and overlapping orthographic representations and a shared semantic representation. A localist connectionist model involving this kind of representation was able to simulate the findings, and also accounted for the cognate facilitation effect found in L2 lexical decision (Dijkstra et al., under revision). Furthermore, the results showed a better performance of the model using a threshold function rather than activation differences between candidates or a Luce choice rule.

Introduction

When learning a second language, one often notices similarities between the first and second language vocabulary. For example *"My name is Bond, James Bond"* is understandable for Dutch language users even in the early stages of learning English. This is because of the resemblance of this sentence with its Dutch equivalent, *"Mijn naam is Bond, James Bond"*. *"Naam"* and *"name"* have clear phonological and orthographic overlap, while *"is"* and *"is"* are form-identical. Because of these resemblances, these English words are easier to comprehend than other, English-specific words.

In psycholinguistics, translation equivalents like *"name"* and *"naam"* that show strong form resemblance between languages are referred to as cognates. Another example is the English word *"night"* that is a cognate with the Dutch word *"nacht"*. Linguistically speaking, cognates are often derived from common roots: *"night"* and *"nacht"* are derived from the Proto-Indo-European word *"*nók$_w$ts"*. Many cognates are words that are borrowed from another language (loan words). Word forms that share their word form, but not meaning, across languages, such as *"leg"* in English and *"leg"* (lay) in Dutch are referred to as interlingual homographs or false friends. In this thesis, we will focus on cognates rather than on interlingual homographs.

It has been argued that, because of the resemblances in cognate-readings, the different readings of the cognate affect each other during processing. As such, they can therefore serve as an interesting predictor of recognition behavior in bilinguals. The recognition of cognates is affected by a number of factors, for instance, their frequency of usage, the number of languages that have readings of the cognate, and the task in which the cognates are processed (Dijkstra, n.d.)(Friel & Kennison, 2001). Importantly, cognates benefit from a larger degree of orthographic and phonological similarity of the readings in lexical decision tasks. It is easier to recognize a letter string as a word when it shares many orthographic features with a word from another language. For example a word like *"night"*, that is similar to its translation equivalent in Dutch, is faster recognized as a word by Dutch-English bilinguals than a word like *"queen"* (which is *"koningin"* in Dutch). The faster recognition speed of cognates relative to matched one-language control words is referred to as the cognate facilitation effect.

What does the cognate facilitation effect teach us about the lexical representations for the cognates in different languages? It has been suggested that cognates share (part of) their representation, leading to a co-activation of both readings during bilingual visual word recognition. In fact, the representation and retrieval process of cognates have been an important domain of bilingual research. Cognates have been used investigate whether there is a shared mental lexicon, and whether this shared lexicon is accessed in a language non-specific manner.

As a first theoretical option, it has been proposed that cognate representations in the bilingual lexicon have linked word form representations (figure 1). Empirical evidence supporting this view was collected by De Groot and Nas (De Groot & Nas, 1991). In their study, Dutch-English bilinguals performed four experiments, comparing within- and between-language repetition-priming (presenting the same word multiple times) and associative priming (presenting a semantically related word before the target) effects. Repetition priming effects were obtained in masked and unmasked conditions for cognates and non-cognates. Associative priming effects were obtained within one language for cognates and non cognates, masked and unmasked, but between languages only for cognates. De Groot and Nas concluded that there may be separate but connected lexical representations for translation equivalents, shared conceptual representations for cognate translations, and separate conceptual representations for non-cognate translations.

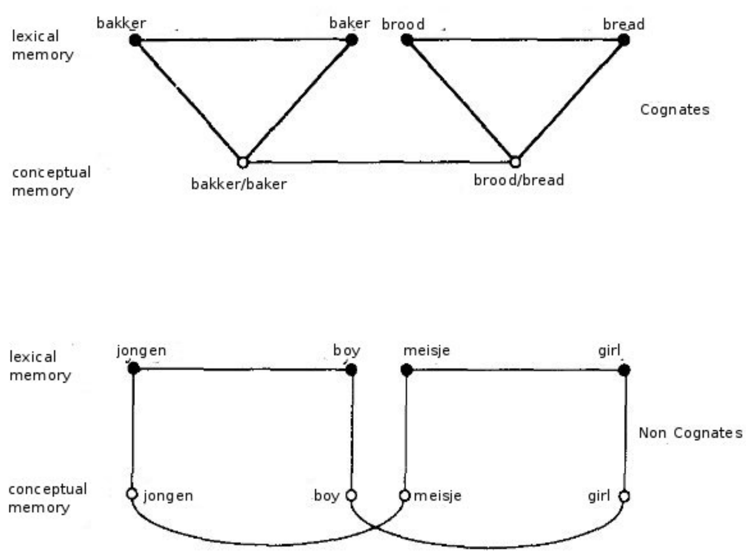

Figure 1. Representation of cognates and non cognates with associative links according to De Groot and Nas (1991)

A second theoretical possibility for cognate representation is provided by Sanchez-

Casas and Garcia-Albea (Sánchez-Casas & García-Albea, 2005). These authors suggest that there is a critical role of morphology in the representation of cognates and non-cognates in the bilingual lexicon. They propose to represent the two readings of the cognate jointly on the basis of a special morphological relation between the two. Reviewing studies with Spanish-English and Catalan-Spanish bilinguals in lexical decision tasks and priming paradigms, they concluded that facilitation effects are only obtained for cognate representations, which combine orthographic and semantic overlap. According to Sanchez-Casas and Garcia-Albea, the separate influence of form similarity is not enough to account for the cognate results, because there are facilitatory effects in non-cognate readings that do not have orthographic or phonological overlap but do have semantic overlap (e.g. semantic priming). Furthermore, there is an absence of facilitatory effects for false friends, which share orthographic overlap, but not meaning.

Moreover, the separate influence of semantics is not enough explain cognate recognition results, because L2 members of pairs of translation equivalents that are non-identical in meaning showed slower recognition times than L2 members of pairs that were identical in meaning. Also, cognates showed priming effects in a semantic priming experiment for all priming conditions (prime duration of 30, 60 en 250 ms) in contrast to false friends and non cognates. Cognate facilitation effects did not differ from effects with morphologically related words within and between languages. The authors interpreted these findings as evidence that the representation of cognates must be joined on a morphological level.

In addition, Duyck et al. (Duyck, van Assche, Drieghe, & Hartsuiker, 2007) investigated cognate processing in a sentence context, to investigate whether cognate effects are also obtained when not presented in separation. Dutch-English bilinguals performed an L2 lexical decision task in which cognates presented in a sentence context cognates (mean 555ms) were recognized more quickly than control words in a similar context (mean 592ms), which interacted with the degree of orthographic overlap ($F(1,31)=4.31$, $p=.054$) meaning that identical cognates were recognized faster than non identical cognates. from which the authors concluded that identical and non-identical cognates are recognized faster in isolation than control words. In a second experiment, the same cognates and control words were used, but where presented as final words in a low-constraint sentence context that could contain both cognate and control word (e.g. *"Lucia went to the market and returned with a beautiful cat [cognate] / bag [control]"*) Again, cognates (mean 632ms) were recognized faster than control words (mean 706ms) in a linear fashion and again an interaction with degree of orthographic overlap was found (F(1,31)=7.88, p=.009). In a third experiment, an eye-tracking experiment, Duyck et al. (2007) found that identical cognates yield shorter read-times than control words. However, for non-identical cognates, there were no significant facilitation or inhibitory effects at all compared to control words. Duyck et al. (2007) concluded that sentence-context may influence cross-lingual interactions early in the recognition process. This provides additional evidence for the non-selective access hypothesis, because again, cognate readings benefit from their cross-lingual counterparts.

In contrast to Duyck et al. (2007) who found a linear effect of orthographic similarity in a lexical decision task, Dijkstra, Brummelhuis, and Baayen (Dijkstra, Brummelhuis, & Baayen, n.d.) found a non-linear effect of orthographic similarity. Dutch-English bilinguals performed an English (L2) lexical decision task in which they processed cognates with a varying orthographic-phonological overlap (e.g. *"lamp-lamp", "flood-vloed"*). Dijkstra, Brummelhuis, and Baayen (under revision) found that orthographically identical cognates were recognized fastest, with a non-linear increase in reaction time for orthographic overlap. There was no facilitation or other effect in reaction time for non cognates. In a second experiment, participants performd a progressive demasking task. In this task, participants had to press a button as soon as they recognized the word that slowly appeared out of a checker box pattern and subsequently they had to type the word they just saw. The reaction times of the progressive demasking data were not dependent on orthographic similarity, but only on word frequency and semantic similarity ratings.

The different findings of these two experiments (facilitation in lexical decision and no facilitation in progressive demasking) were interpreted as evidence for a difference in task demands. The authors concluded that to account for the results of the second experiment, a shared lexical representation for cognates is not required. In a shared orthographic representation for cognate readings, the progessive demasking task should show cognate facilitation, because both readings of the cognate become activated when their orthographic form is shared. The results of both experiments can be explained by assuming that form overlap between two readings of a cognate leads to frequency-dependent parallel activation of two form-representations of a cognate that activate a (partially) shared semantic representation (figure 2).

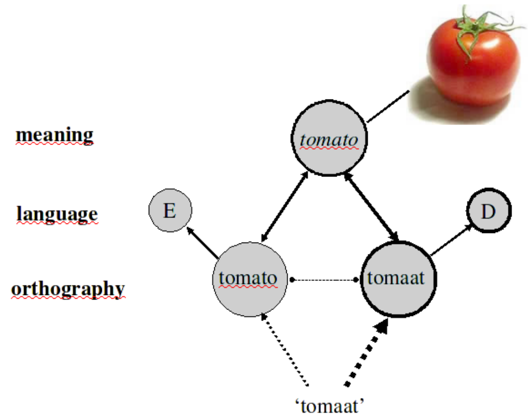

Figure 2. Orthographic-Semantic Representation of Cognates

Besides the theories about the representation of the cognates, there have also been hypotheses about the process of lexical access of the words. Most importantly, the cognate facilitation effect has lead to believe that cognates are accessed in a language non-specific fashion; e.g., the readings in both languages affect each other, so they can not be fully separated. This is called the non-selective access hypothesis, if this hypothesis is valid, beside cognate facilitation effects for cognates existing in two languages, additional effects should be found for triple cognates (cognates existing in three languages).

Lemhöfer, Dijkstra, and Michel (Lemhöfer, Dijkstra, & Michel, 2003) investigated whether the language non-selective access hypothesis also holds for processing cognates that exist in three languages. Dutch-English-German trilinguals performed a German lexical-decision task in which they had to decide whether a presented letter string was an existing German word or not. The following materials were used: double cognates (translation equivalents overlapping in form for Dutch and German, not in English such as *"kunst", "Kunst"* and *"art"*), and triple cognates (translation equivalents overlapping in all three languages such as *"naam"*, *"Name"* and *"name")*, German control words and non-words. Double cognates were processed faster than control words (634 ms 688 ms = 54 ms, mean), and triple cognates were processed even faster (610 ms 688 ms = 78 ms, mean). A monolingual German control group did not show these differences in reaction time for cognates and controls, indicating that the materials were well-matched across conditions. The authors interpreted the facilitation results of their experiment as evidence for the non-selective access view, implying that all languages known by an individual affect word activation (and thereby word recognition).

Another explanation of the cognate effects found is that the facilitation effects for cognates in lexical decision as opposed to no facilitation effects for cognates in progressive demasking are cause by a difference in task demands. Cognates may be unable to benefit from the similarity between the cognate readings, when the task in which the cognate is processed has heigher demands, which explains why no facilitation effects are found in progressive demasking (this task may be more demanding than lexical decision).

Therefore, Font (Font, 2001) investigated the influence of task demands by performing a French-Spanish lexical decision task in which he found the cognate facilitation effect. She also performed a French-Spanish language decision task, e.g., participants were shown French and Spanish words such as *"livre"* and *"libro"*, they had to decide whether the word was French or Spanish by pressing the corresponding button. In this task, language-information of the word had to be activated. She concluded that instead of a cognate facilitation effect a cognate inhibition effect occurred. This shows that lexical candidates from both languages are activated, again providing evidence for a non-selective access hypothesis and shared semantic representation as suggested by Dijkstra, Brummelhuis, and Baayen (under revision).

The results of Lemhöfer et al. (2004) can easily be explained by the suggestion of a (partially) shared semantic representation for cognates as suggested by Brummelhuis, and Baayen (under revision). This representation also allows the possibility for task dependent effects. It has been shown for false friends that such effects exist, but for cognate effects the research is limited.

In the present study we tested the hypothesis of task-dependent cognate effect by means of a Dutch-English language decision task. In extension we investigated whether the proficiency in second language also affected the cognate effects.

First of all, the language decision task may lead to a cognate-inhibition effect such as Font (unpublished) found because of the competition that rises between words that are highly similar. A language decision task can easily be translated into a lexical decision task in which a participant replaces the question. *Is this word English or Dutch?* with *Is this word English or not?* which shows that the tasks are not that different. Nevertheless, Font showed that the results are very different. This shows the importance of task demands. These findings have led us to expect an inhibition effect of orthographic similarity in the language decision task in stead of the facilitation effect found in lexical decision, due to competition between words that are more similar. This effect is expected to be non-linear because of recent findings in the lexical decision task (Dijkstra et al., n.d.).

Furthermore, because cognate effects can also be influenced by the participants L2 proficiency, e.g. the better you master your second language, the more the second language can compete with the first language, I will also investigate the influence of proficiency on the cognate effects. And in light of the proficiency I will also investigate if the effects are the same for words from the first language (L1) and words from the second language (L2). We expect that higher proficient participants will be faster in recognizing the words from their second language. Furthermore, it is possible that high proficient participants receive more competition from the L2 readings of the translation equivalents which will result in an interaction between proficiency and orthographic similarity of translation pairs.

Finally, If there is indeed a non-linear orthographic similarity effect as found by Dijkstra et al.(under revision), there are several possible explanations. The effects may be influenced by the decision criterion used by the participants, e.g. on what strategy do bilinguals decide what language the word belongs to, or the effect is caused by the representation of the cognate. This will be investigated by comparing several decision criteria in an implementation of a model on bilingual word recognition, on which I will explain more in the second part of this article.

Experiment

The first part of this research consists of an experiment to investigate the influence of task demands, proficiency, language and orthographic similarity on the recognition of cognates.

The results will be compared to normal translation equivalents that share less orthographic features and are not rated as being cognates in the rating task performed in Dijkstra, Brummelhuis, and Baayen (under revision).

Method

Participants. Twenty-four subjects (mean age 23.3 years, 17 women, 7 men) took part in this experiment. They were all students of Radboud University Nijmegen. They all had Dutch as their native language and had experience with the English language for at least 8 years. English was often used in their study, and all participant read English on the internet or watched English television regularly. Twelve of the participants had above average experience because they were third year or higher students of English, had stayed for more than six months in an English-speaking country or had English family. All participants were paid for their participation or received course credit. They were asked to fill out a questionnaire about their experience in English, which resulted in the following data.

Stimulus Materials. As a list of test word pairs with a variable degree of cross-linguistic orthographic and phonological overlap (from identity to no overlap whatsoever), we used the words rated in the experiment of Dijkstra et al. (under revision) as a starting point. This was a list of 360 word pairs containing cognate-pairs and non-cognate-pairs as control words to be used. However, some of the words were not suited for the language decision task, because of multiple readings in both languages (e.g. the Dutch form of the pair *"bath-bad"*, also has an English reading, because there is no context *"bad"* can be read in English or in Dutch, although we intended the Dutch reading). Therefore, two cognate-pairs were removed from the set of words from Dijkstra et al. (under revision). Furthermore, 16 control-pairs were replaced by new word pairs that did not have multiple multilingual readings. The new pairs were matched on word frequency and word length and had maximally one overlapping letter. See Appendix A for the complete list of stimuli.

Furthermore, the frequencies of English words were calculated and the word form frequencies of Dutch translations were added on the basis of the sum of the instances of word form frequencies in the Celex database (Baayen, Piepenbrock, & Gulikers, 1995). For

Table 1: Results of the questionnaire on English experience

	Proficiency	
	Low	High
Age	22.17	22.75
Year of Study	2.83	3.73
Age of Acquisition L2	11	9.42
Duration of L2 experience	10.75	13.58
Frequency of reading English literature[a]	1.92	1.25
Frequency of reading English school literature[a]	1.5	1.25
Frequency of writing English[a]	3.42	1.75
Frequency of speaking English[a]	3.08	2.25
Frequency of watching English television[a]	2.08	1.75
English reading experience[b]	4.58	5.42
English writing experience[b]	3.67	5.08
English speaking experience[b]	3.75	5.67

[a] with 1 being very often, and 4 begin never.

[b] with 1 being very little experience and 7 very much experience (in comparison to other students)

the Dutch words these frequencies were based on a Mln-calculation of the INL-corpus and for the English words the frequencies were based on a Mln-calculation of the COBUILD-corpus in Celex.

Procedure. Participants were seated in a lit room at a distance of 50 cm in front of a computer screen. Before the experiment started, the participants were asked to complete a checklist that measured their experience with English.

The stimuli were presented one by one on the screen of a PowerMac 3.6 computer using Mac OS 9.2. The stimuli were selected from four balanced list of cognate and non-cognate control words. Each list contained 158 English and 158 Dutch words and was balanced according to frequency and cognate similarity e.g each list contained approximately the same number of high and low frequency words and high and low similarity in cognate pairs. In total, each participant processed 330 stimuli in a unique semi-random order, spread over 4 blocks, with one test-block of 14 items for the participants to become familiar with the experiment. The participants were asked to decide whether the presented word was a Dutch or an English word. This was done by pressing the right or the left button. This response button allocation was varied over participants to avoid hand dominance issues. Each participant only saw one of the readings of a translation pair. Half of the presented items consisted of the English reading, and the other half of the Dutch readings. Furthermore, they were asked to make a decision even if they did not know the answer. If a stimulus was a possible word form in both languages, they were told to respond to the reading that came to their mind first.

Results

The data was statistically analyzed using R version 2.6.1 and 2.8.1 (R Development Core Team, 2007, 2008)(Baayen, 2008) (Baayen, Davidson, & Bates, 2008). There were no participants or items that were excluded from the dataset, because none of them had high error rates (above 15%). Data points below 300 ms were removed, because these data points are likely to be artifacts.

Inspection of response latencies showed a non-normality. A comparison of log transform and inverse transform showed that the inverse transform was successful in attenuating the non-normality. All reaction times were transformed using RT = -1000/RT to ensure a positive correlation between original and transformed RT's.

The data was analyzed using a linear mixed effects model, only correct responses were analyzed. Subjects and words were marked as cross random effects. The data was fitted to the model, after which outliers (data points with residuals exceeding 2.5 standard deviation points) were removed, resulting in a dataset consisting of 7133 points. A new model was fitted on this dataset using the same predictors.

The predictors used in the main linear model were reaction time on previous trial, trial number, orthographic rating, frequency per million and language of the word. Subjects and words were taken as random intercepts. As in Dijkstra, Brummelhuis and Baayen (under revision), orthographic similarity emerged as a non-linear predictor.

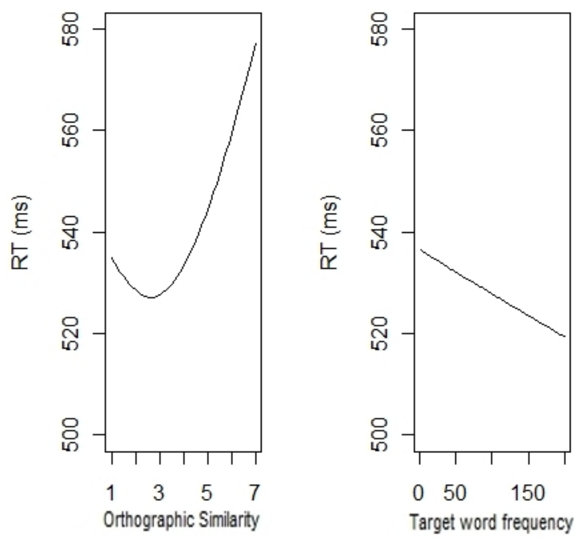

Figure 3. Influential predictors in the Language Decision Task

Figure 3 visualizes the effects of the predictors of the model and table 2 the corresponding model estimates. The data revealed several significant effects. In table 3 you will find the corresponding ANOVA table. As expected there is an overall effect of frequency (p=0.0316), higher frequency words lead to faster responses. This is consistent with the frequency effect found in other research. There is no significant interaction between frequency and language.

Additionally, we found a nearly significant effect of target word language (p=0.0564). English words are responded to faster than Dutch words. The strongest effects in the model are those of reaction time in previous trial, reaction time over trial numbers and orthographic rating of the word (corresponding to cognate status). This shows that participants become slower at the end of the experiment (high trial numbers), and that participants the were slow on the previous trial are likely to be slow on the present trial, and previously fast participants are likely to respond fast again. The effect of orthographic similarity is non-linear and is opposite to the facilitatory effects found lexical decision tasks. In the language decision task they are reversed, and it shows that for cognates, the higher the overlap between words, the slower the recognition to that word ($p < .0001$).

Table 2: A linear mixed model for the Language Decision Task

	Estimate	Std. Error	HPD95lower	HPD95upper	t-value
(Intercept)	-1.4534	0.0427	-1.5366	-1.4757	-34.07
PrevRT	0.1926	0.0114	0.1716	0.2159	16.92
O.rating (Linear)	-0.0261	0.0089	–0.0419	-0.0095	-2.91
O.rating (Quadratic)	0.1253	0.0253	0.0810	0.1716	4.96
Frequency	-0.0003	0.0001	-0.0005	-0.0001	-2.43
Target Language	-0.0444	0.0203	-0.0811	-0.0060	-2.19
Frequency * Language	0.0003	0.0002	0.0000	0.0006	1.75

In a second model fitting, years of experience in English (proficiency) was added as predictor for the model. Figure 4 visualizes the effect of proficiency in English as predictor for the language decision task. Although proficiency itself was not significant as predictor (p=0.77), the interaction between proficiency and target language was significant (p=0.04), which means that the longer the experience with English, the faster the reaction time will be on English target words. It also seems that participants with higher proficiency in English, also seem a little faster in recognizing Dutch words.

Table 3: ANOVA: Language Decision Task

	Df	Partial SS	MS	F	P
PrevRT	1	143.9702245	143.9702245	1119.65	$< .0001$
Orthographic Similarity	2	17.7279184	8.8639592	68.93	$< .0001$
Nonlinear	1	5.5911872	5.5911872	43.48	$< .0001$
Target Language	2	0.7398175	0.3699088	2.88	0.0564
Target Frequency	2	0.8891128	0.4445564	3.46	0.0316
REGRESSION	6	161.2348197	26.8724699	208.99	$< .0001$
ERROR	6827	877.8498853	0.1285850		

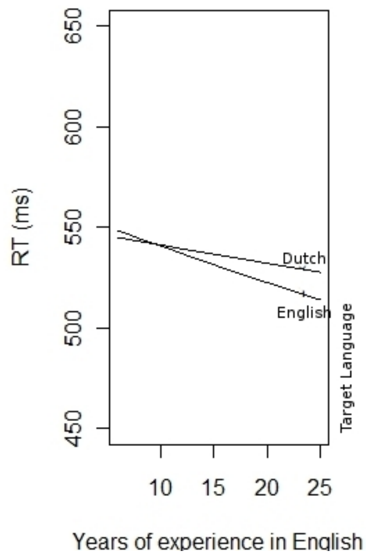

Figure 4. Effect of proficiency in the Language Decision Task

Discussion

When cross-linguistic orthographic similarity between Dutch en English translation equivalents increased, the reaction times in the language decision task became slower. This effect was the largest for identical cognates, which can be explained by the large competition between the languages because both reactions are allowed. There was an increase of 34 ms for cognate readings. For targets that only partially overlapped, a small facilitatory effect was found (11 ms). The distribution of errors shows that there was not a significant increase in error with increase in orthographic similarity. Although, as figure 5 shows, there is an outlier in error rate for nearly identical cognates.

The idea that cognates share a (morphological) representation in the lexicon suggests that using a formulated task account, inhibition effect such as found in the present study, would depend on word frequency and not on cross-linguistic similarity. Although there are effects of frequency, the effects of orthography are not in line with the idea of a shared morphological representation.

The associatively linked cognate theory has trouble predicting separate inhibition an facilitation effects in cognates. The theory does not predict inhibition effects as a consequence of selection competition.

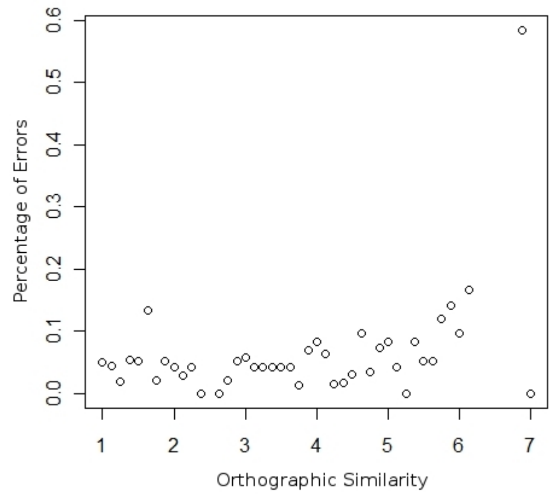

Figure 5. Influence of Orthographic Similarity on Error Rating

The non-linearity of the orthographic similarity effect does not fit well with distributed connectionist models, and fits better with a localist connectionist model as suggested by Dijkstra, Brummelhuis and Baayen. They suggested that these cognate recognition effects van be understood in terms of an activation of two separate orthographic representations, a (partial) overlapping semantic representation. This idea is extended with single representations for lexical properties such as language. This will be further investigated in the second part of the study, which will involve modeling the data collected in the language decision task.

Simulations

(In cooperation with S. Rekké)

An efficient way for testing theories about the bilingual brain is to make models of them. By running data through these models additional evidence can be obtained for the theories underlying the model. In light of the results of the experiment above I will introduce a few models concerned with the bilingual lexicon.

On the basis of the various studies on cognate processing, Dijkstra and van Heuven

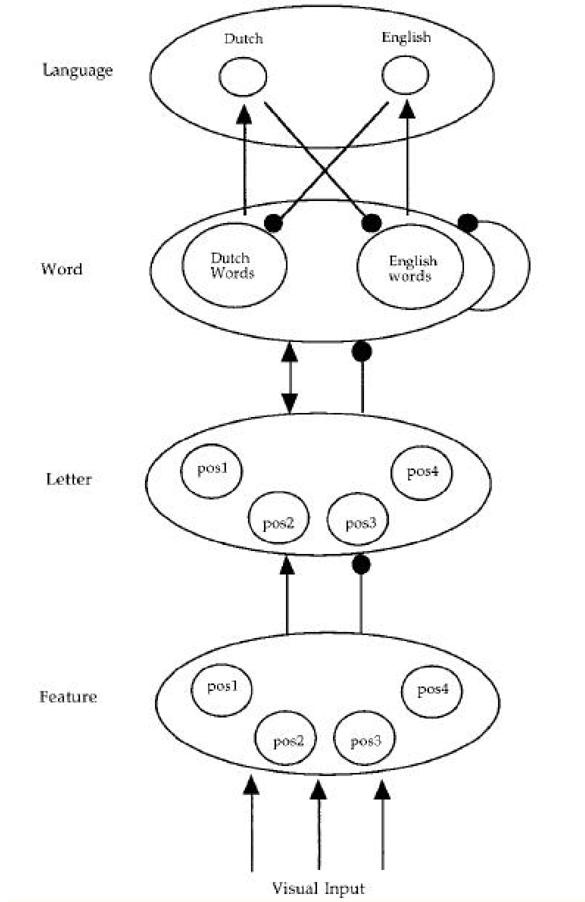

Figure 6. Bia-model

(1998) (van Heuven & Dijkstra, 1998) propose the BIA-model for bilingual word recognition as illustrated in Figure 6. This computational model of bilingual word recognition is based on the Interactive Activation model of McClelland and Rumelhart (1981, 1982, 1988)(McClelland & Rumelhart, 1981)(McClelland & Rumelhart, 1982)(McClelland & Rumelhart, 1988), which implements bottom-up word activation in a language non-selective fashion (letters activate words from all languages). Language-nodes are added that serve as linguistic representations for language membership (which language does the word belong to) and linguistic functional mechanisms (they collect activation from all lexical representations within a language). The BIA-model is limited by the lack of phonological and semantic representations and the underspecification of representation of homographs and cognates. Furthermore, there is only a limited account of the effect of context.

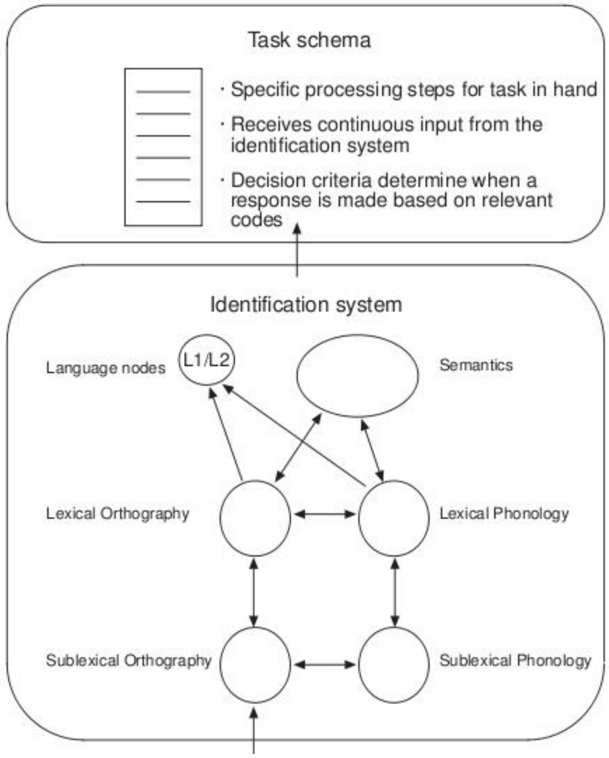

Figure 7. Bia+-model

Therefore, in light of new empirical evidence on bilingual word recognition, Dijkstra and van Heuven (2002) (Dijkstra & van Heuven, 2002a) (Dijkstra & van Heuven, 2002b) proposed an extension to the BIA model, referred to as the BIA+ model. This model is displayed in Figure 7. The earlier model was extended by adding phonological and semantic lexical representations to the orthographic nodes. In addition, the task of the language node was restricted to its representational function. Furthermore, a task-decision system was introduced to distinguish the effects of a non-linguistic (instruction, stimulus-list) and a linguistic (sentences, discourse) context.

In addition Rekké (2009) created a new model based on the IA model and BIA model and using the java framework (see figure 8). Furthermore, the model uses the cognate representation strategy as provided by Dijkstra, Brummelhuis and Baayen (submitted) (Dijkstra et al., n.d.). The model includes a separate task and decision layer but does not have a sub-lexical level. It allows for orthographic and transcribed phonological input. By creating a network of semantic connections as found in the Nelson database, the model allows for semantic interaction. Language nodes are involved in the activation process. These nodes inhibit all word nodes of other languages and facilitate words of their own language, this allows for a so-called language-mode, in which words of a specific language are favored. In extension to the ideas of the BIA+ model, the model by Rekké can simulate perception (semantic priming, language decision, lexical decision) and production data (translation), and is very versatile as the model allows for multiple language lexicons.

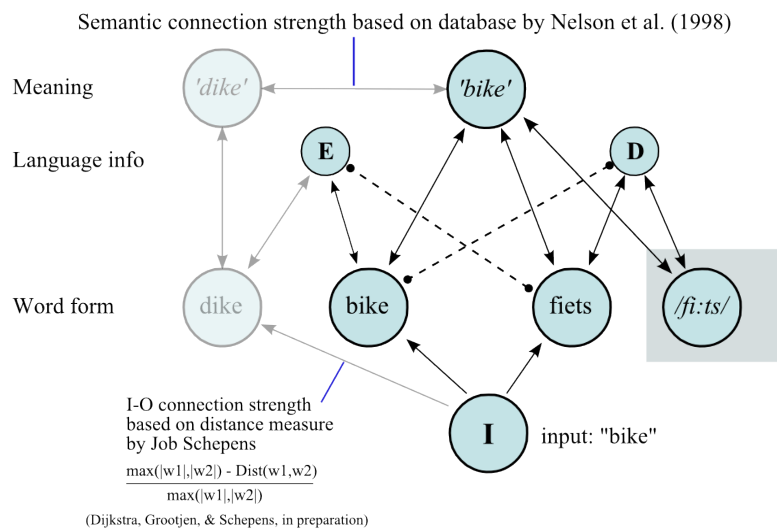

Figure 8. Word Translation Model by Steven Rekké

In the present study, Rekké and I will be cooperating by comparing data provided by the model to the language decision data provided by M. Sappelli and lexical decision data provided by Dijkstra, Brummelhuis and Baayen (Dijkstra et al., n.d.). We will compare these sets and investigate whether the output from the model is similar to the language and lexical decision data. We expect that the model will have no problems in simulating the language and lexical decision data. Furthermore, we will investigate what the influence of decision criteria is on the fitness of the model data on the datasets. We expect that a simple threshold function is enough for simulating the different effects and that the correlation to the empirical data is largely dependent on the parameter settings used in the model.

Method

Parameter Settings. The model mainly uses the parameters used by the IA and BIA model. Because the lexical access in the model by Rekké is different from the access in the (B)IA model, in the sense that there is no sub-lexical layer, there is a new IO-parameter which was manually determined, to provide all sorts of common lexical effects. Furthermore, we have tried several settings and the final model parameters do not include inhibition from orthographic nodes to opposite language nodes, but do include strong facilitation to language nodes and weak facilitation and normal inhibition from language nodes to orthographic nodes.

Tasks. We tested the performance of the model on two tasks, e.g. lexical decision and language decision. In the lexical decision task, the model gives a word and time step as output that corresponds to the word that is recognized. The decision is based on the word that received the most activation, the time step is returned on the basis of the decision criteria selected. English lexical decision can be seen as the activation of all English words compared to the activation of all other words, general lexical decision can be seen as the activation of the entire network. For the language decision task, the model returns a language and time step as output. The decision is based on the activation value of the language node that has the highest activation in combination with the decision criteria selected.

Testdata. Because the model was not fit on specific words, there was no separate train or test set. The parameters are not fit on correlation to reaction times of the experiments and were only fit on effect-existence. Therefore, for the language decision task, the model was tested on the entire word set as used in the experiment. Because the model does not incorporate repetition or semantic priming effects in the setting we used for the test, all words were run through the model to obtain as much data as possible. For the (English) lexical decision task, only the English words of the test list were presented. No non-words were shown, because we are only interested in the effects on words, but it is possible to decide on non-words.

Decision Criteria. The dependent variable in the research is the decision criterion. This is the parameter setting that is varied over the simulations. For each task, we have run the model three times, using three different decision criteria. These are, a simple threshold function in which the recognized word will be the word that reaches the set threshold of 0.7 first (referred to as threshold criterion). The second criterion was the function in which the highest activation is compared to the activation of the second highest, and when the difference between the two is more than 0.7, the word is recognized (referred to as 1-2 difference). As final criterion we used a Luce choice rule, which calculates the ratio between highest activation and the activation of the whole net, and when this ratio is above 0.7 the word is recognized (Luce choice criterion).

The thresholds used are either based on literature (0.7, used in IA-model of McClelland & Rumelhart, 1981,1989 (McClelland & Rumelhart, 1981)(McClelland & Rumelhart, 1982)(McClelland & Rumelhart, 1988)) or based on own experience (also 0.7).

Results

Language Decision. Figure 9 shows the predictions the model made for a language decision task. Only the first criterion, the threshold function 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 significance 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 remarkable high error rate of 57%. The 1-2 difference did not predict non-linear effects of orthographic similarity, in fact, there was no significant influence of orthographic similarity at all (p=0.49). The criterion did show a significant effect of 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 threshold criteria and first-second criteria yielded exactly the same results, because inherently the two are the same using these parameters. This is because there are only 2 active nodes (only two nodes are involved) of which only one reaches an activation higher than 0.0, resulting in the same situation as in the threshold situation. 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 course of the effect seems to be different as can be seen in figure 10. Furthermore, there is no significant frequency effect found by the Luce choice ratio, although the course seems to be the same as the other criteria. The correlation of the threshold and difference 1-2 criteria with the experimental data is 0.197, based on 482 data points. This is much higher than the correlation found without orthographic-language inhibition.

Furthermore the model is able to make mistakes. On the language decision data with orthographic-language inhibition, using the threshold or 1-2 criterion, the model has an error-rate of 8.7% (46 errors). Also, of these errors 35% resulted in no response at all, because of too much competition between the choices. 50% of these no-response cases included identical cognates. Of the remaining errors, more faults were 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 which 90% were identical cognates and the remaining 10% were nearly identical cognates (1 letter difference). Of the remaining errors 74% is cognate, and using this criterion the model also predicts more mistakes on English than on Dutch words (63%)

Interestingly, the model without O-L inhibition shows less 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). It seems that allowing O-L inhibition gives a better overall performance.

Lexical Decision. Using the model with O-L inhibition, there were no clear lexical decision effects found at al. Therefore, we continued the simulations using the model without O-L inhibition, although these results were slightly less good for the language decision task.

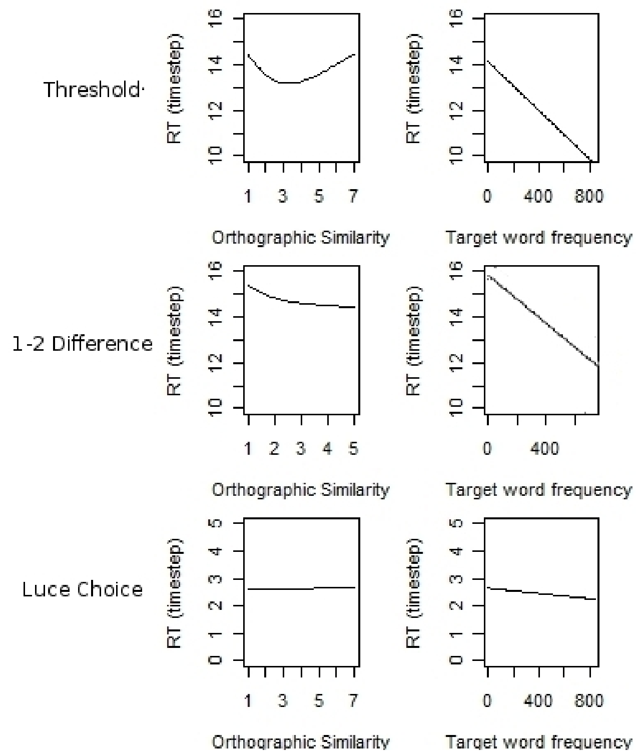

Figure 9. Modeling Language Decision without Orthography-Language inhibition

In the lexical decision task we found strongly significant non-linear effect of orthography ($p < 0.0001$), which reveals a facilitatory effect of similarity (figure 11) like the one Dijkstra et al.(under revision) found (figure 12). 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 one that predicted a significant frequency effect ($p < 0.0001$), the effect predicted is comparable to the effect Dijkstra et al. (Dijkstra et al., n.d.) found (figure 11). No significant interaction effect with target language was found (p=0.1).

Error-rates in the lexical decision task were very low (0% for threshold) or very high (73% for 1-2 difference and 47% for Luce choice), which shows that only the threshold criterion seems to be a good function for deciding on visual word recognition.

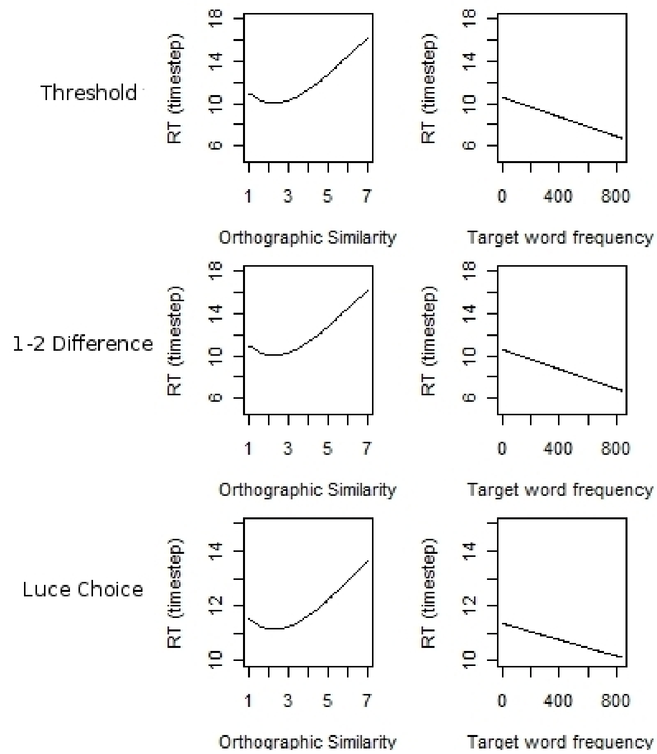

Figure 10. Modeling Language Decision with Orthography-Language inhibition

Discussion

As expected a local connectionist model is efficient in modeling the non-linear effect of orthographic similarity between Dutch and English words. It seems that there is a clear influence of decision criterion in a lexical decision task, but in a language decision task the different criteria are translatable into each other, dependent on the parameters used. The Luce choice rule is definitively not a good criterion for recognition behavior because of its artificial results and low significant predictability. Furthermore, the 1-2 difference criterion is not a good criterion because of its high error-rates. We did find that the 1-2 different criterion can be used for providing a recognition time step of non-words (in stead of non-words being words that are not recognized within the maximum of time steps) although it is probably best to let another criterion decide whether it is a non-word or not.

These results show the importance of separate task and decision layers. Cognate effects are highly dependent on the task and this means that although the basic system

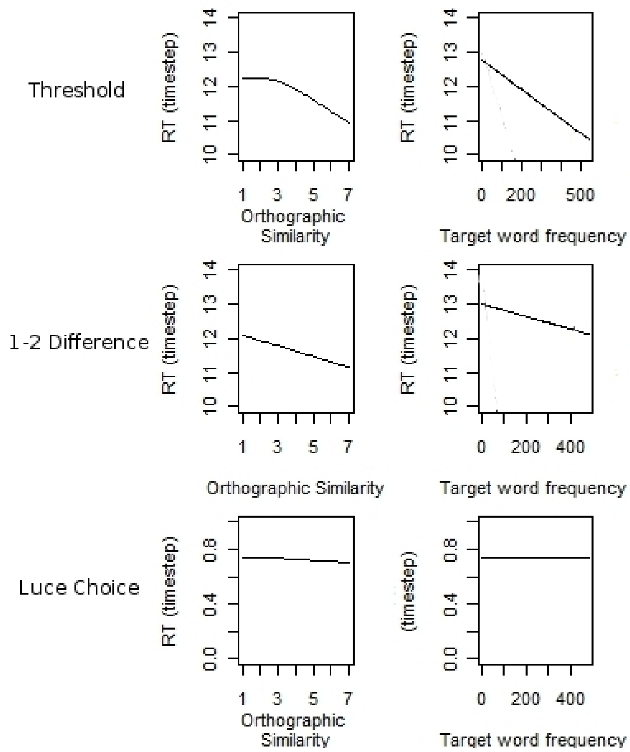

Figure 11. Modeling Lexical Decision without Orthography-Language inhibition

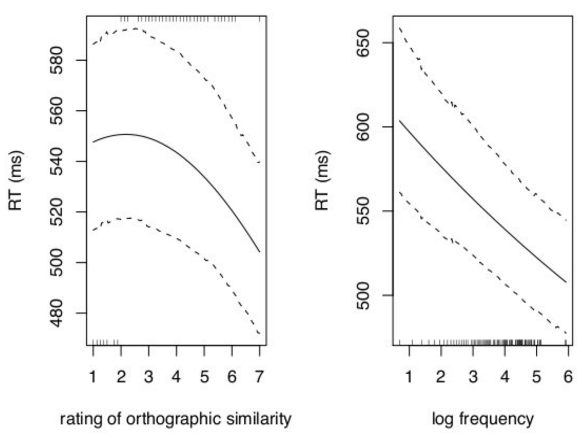

Figure 12. Lexical Decision effects in study of Dijkstra et al. (under revision)

may be the same, there is a big influence of the actual task. Decision criterion performance is also highly dependent on the underlying system and parameters.

General Discussion

In this paper, we investigated cognate processing by means of a language decision experiment and by modeling the data in a localist connectionist model. Furthermore, we investigated how different decision criteria affected the results of the model.

In the language decision task, we obtained a non-linear inhibition effect of orthographic similarity for cognates on RTs. An increasing inhibitory effect was found for word pairs with similarity ratings larger than 3 (on a scale from 1 to 7). A low rated word as *"kleur-colour"* (rated 2.00) has a mean recognition time of approximately 520 ms and a high rated word as *"debate-debat"* (rated 6.00) has a mean recognition time of approximately 548 ms. This effect is opposite to the facilitation effect of orthographic similarity found in lexical decision (Dijkstra et al., n.d.).

The non-linear orthographic similarity effect found in the language decision task is in contradiction to the linear effect found by Font (unpublished). However, the non-linearity is also found in the lexical decision task by Dijkstra, Brummelhuis and Baayen (under revision). It is possible that the difference in findings can be explained by a more binary distinction of Font (cognate vs non cognate) and a more continuous distinction by Dijkstra, Brummelhuis and Baayen (under revision) and the present study (orthographic similarity rated on a scale from 1 to 7). The found overall frequency effect is in line with previous research (Dijkstra et al., n.d.)(Font, 2001).

Besides the results obtained from the language deicision experiment, we showed that a localist connectionist model (created by Steven Rekké) with a separate task layer is able to simulate the observed inhibitory effects on RTs in the language decision task, as well as the facilitatory effects on RT in the lexical decision task. It was even possible to use the same parameter settings to model both tasks, although it may perhaps be better when different parameter settings are allowed for each task, because this yield to stronger simulation results, more similar to empirical data.

Furthermore, we compared simulation results for both tasks with respect to three decision criteria: activation threshold, activation difference between most activated and second activated candidate, and Luce choice rule. In our language decision experiment, the activation threshold and activation difference criteria led to exactly the same results when inhibition links from orthography to language were implemented. Otherwise, the activation difference criterion yielded a poor fit to the empirical data. Finally, the Luce choice ratio led to artificial and non-significant results that were not comparable to the participants performance.

With respect to the simulations of the lexical decision task using the activation threshold criterion, the results showed a facilitatory effect of orthographic overlap on processing time. The observed effect was non-linear and comparable to the empirical effect observed by Dijkstra et al. (under revision). However, the other decision criteria predicted linear effects on processing or no effects at all. Only the threshold criterion predicted a significant effect of word frequency that was comparable to that reported by Dijkstra et al. (under revision).

Our empircal language decision results can be interpreted in terms of the different representational and processing accounts discussed in the Introduction. Although these accounts were often only verbal and incomplete, we derived reasonable predictions of them to account for cognate processing in the language decision task.

The first account was a morphologically shared representation for cognates. We did not find significant support for a morphological representation account. This account predicts frequency dependent increase or decrease in reaction times for more identical cognates (because it assumes shares morphological representations across languages), but in contrast we found no interaction between frequency and orthographic similarity.

According to the second account, the linked word-form hypothesis, cognates are represented as associatively linked and semantically shared words.This theory predicts the same frequency dependent cognate effects as the morphological theory, but because no interaction was found, we did not find significant support for the linked word-form hypothesis.

However, our results do support the third account, shared semantical representation, as proposed by Dijkstra et al. (Dijkstra et al., n.d.). This account assumes a shared representation on the semantics level, which predicts cognate facilitation or inhibition effects that are not dependent on frequency.

Future research for the language decision experiment includes a more thorough investigation of the non linear effect to gain more understanding where this non-linearity comes from. Additionally, the language decision research can be extended by performing a language decision experiment with more than two languages, to see how this affects the inhibitory cognate effects. Furthermore, the decision strategy of participants can be further explored, e.g. what is the influence of the preceding target, and how well does the strategy participants think they use fit to the actual results.

The model by Rekké can be considered as an implementation of the third account (e.g. shared semantics), in which cognates are represented as two seperate orthographic forms and one shared semantic form as proposed by Dijkstra et al. (under revision) The model was able to simulate the non-linear cognate inhibition effect in language decision as well as the non-linear cognate facilitation effect in lexical decision. The mechanism that allowed the model to do this was task dependence, implemented by means of a seperate task layer. This layer determines which node is the winning node and thus the output of the model. For different tasks, different activation readings were used. For language decision, the simulation results are bases on the highest activated language node, for lexical decision the results were based on the highest activated orthographic form node. This task dependent mechanism is an important aspect of the model by Rekké.

Simulations with different decision criteria showed that the model performed best when the threshold criterion was used. In fact, not every criterion was useful for both tasks. Especially the Luce choice criterion was too artificial, because this criterium predicted only a few discrete cycle times. It predicted that either no recognition occured, or it occured at timestep 2.7, 4.7 or 6,7. Thus, it is not likely to correspond to the decision criterion humans use unless other mechanisms are added. The activation difference criterion did not function very well either for predicting word processing, but might be more suitable for establishing non-words rejection times. Future investigations with respect to this point are recommended.

The model by Rekké is currently able to simulate lexical and language decision, semantic priming and word translation using a shared semantic representation for cognates, but future research is possible. The model could be extended by introducing shared representations for false friends (on an orthographic level) and for ambiguous meanings (multiple semantic representations connected to one orthographic node) as in figure 13. Furthermore, simulation results may be improved by taking a closer look at the parameters used for the different tasks. These are not yet explored to the fullest. It was not in the scope of the research to find the parameters that provides the best fit to the empirical data. We limited our research by using the BIA parameters with minor modifications. It is promising that these parameter settings already yielded such good results, and it is probable that additionial modificiations to the parameter settings may provide an even better fit to the empirical data.

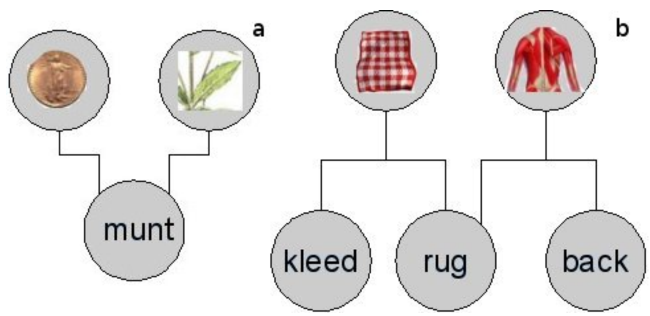

Figure 13. Proposed representations for ambiguous meanings (a) and false friends (b)

Additionally the performance of the model can be compared to the performance of other models. For example, semantic priming and lexical decision simulation can be compared to the simulation by the semantic interactive activation model (van Delft, Sappelli, Dijkstra, in preparation). This model does not inculde the same shared semantic representation as used in the word translation model by Rekké and thus different results are predicted.

To sum up, we found evidence for the language non-selective access hypothesis, because in the language decision experiment, words from different languages influenced each other through response competition. Additionally, we found evidence for the representation of cognates by means of a shared semantic representation, because the response competition in language decision can come from the shared semantic representation that activates both readings of a cognate and therefore the readings are harder to distinguish when the orthographic forms are more alike. Furthermore, the empirical data can be explained in terms of the same underlying mechanism used in lexical decision but with different task demands, which is concluded from the opposite effects in the tasks (cognate inhibition in language decision versus cognate facilitation in lexical decision). And finally, we showed that using the shared semantic representation and task-dependence mechanism in a word translation model it is, among other effects such as word frequency, word length, semantic and orthographic priming, also possible to simulate the task dependent cognate effects found in the empirical data.

References

Baayen, R. (2008). *Analyzing linguistic data. a practical introduction to statistics using r.* Cambridge University Press.

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language.*

Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). *The celex lexical database.* CD-ROM. Philadelphia, PA: Linguistic Data Consortium.

De Groot, A., & Nas, G. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language.*

Dijkstra, A. (n.d.). *Met andere woorden over taal en meertaligheid.*

Dijkstra, A., Brummelhuis, B., & Baayen, R. (n.d.). *Cognate effects in bilingual word recognition.*

Dijkstra, A., & van Heuven, W. (2002a). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition.*

Dijkstra, A., & van Heuven, W. (2002b). Modeling bilingual word recognition: Past, present and future. authors response. *Bilingualism: Language and Cognition.*

Duyck, W., van Assche, E., Drieghe, D., & Hartsuiker, R. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Font, N. (2001). *Rôle de la langue dans l'accès au lexique chez les bilingues: Influence de la proximité orthographique et sémantique interlangue sur la reconnaissance visuelle de mots.* (Unpublished Doctoral Thesis of the Université Paul Valery, Montpellier, France)

Friel, B., & Kennison, S. (2001). Identifying german-english cognates, false cognates, and noncognates: methodological issues and descriptive norms. *Bilingualism: Language and Cognition.*

Lemhöfer, K., Dijkstra, A., & Michel, M. (2003). Three languages, one ECHO: Cognate effects in trilingual word recognition. *Language and Cognitive Processes*.

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*.

McClelland, J. L., & Rumelhart, D. E. (1982). An interactive activation model of context effects in letter perception: Part 2. the contextual enhancement effect and some test ans extensions of the mode. *Psychological Review*.

McClelland, J. L., & Rumelhart, D. E. (1988). *Explorations in the microstructure of cognition: A handbook of models, programs, and exercises*. Cambridge University Press.

Sánchez-Casas, R., & García-Albea, J. (2005). The representation of cognate and noncognate words in bilingual memory: Can cognate status be characterized as a special kind of morphological relation? *Handbook of Bilingualism: Psycholinguistic approaches*.

van Heuven, W., & Dijkstra, A. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and language*.

Appendix

Stimulus Material

On the next pages you will find the stimuli used in the language decision experiment. Table A1 represents the cognate words used in the experiment and simulations, table A2 represents the corresponding control words.

Table A1: Cognate stimuli in the language decision experiment

English Word	English Frequency	Dutch Word	Dutch Frequency	Orthographic Similarity rating
colour	85	kleur	97	2.00
degree	100	graad	13	2.13
salt	44	zout	41	2.13
thumb	24	duim	25	2.13
south	199	zuid	8	2.25
thirst	6	dorst	17	2.25
lion	18	leeuw	15	2.38
heaven	40	hemel	97	2.63
core	17	kern	35	2.63
screen	30	scherm	15	2.75
flood	14	vloed	8	2.75
love	367	liefde	168	2.88
hour	162	uur	605	2.88
seed	29	zaad	21	2.88
oath	6	eed	8	2.88
rich	124	rijk	92	3.00
grey	86	grijs	26	3.00
thin	78	dun	20	3.00
youth	64	jeugd	63	3.00
soap	21	zeep	16	3.00
foot	104	voet	96	3.13
cellar	11	kelder	22	3.13
needle	10	naald	11	3.13
thorn	5	doorn	7	3.13
nose	76	neus	98	3.25
honey	21	honing	12	3.25
fist	19	vuist	23	3.38
mill	10	molen	8	3.38
strong	170	sterk	208	3.50
sugar	56	suiker	39	3.50
guide	40	gids	19	3.50

Table A1: (Continued)

English Word	English Frequency	Dutch Word	Dutch Frequency	Orthographic Similarity rating
mouse	9	muis	9	3.50
month	90	maand	74	3.63
tower	49	toren	21	3.63
anchor	5	anker	8	3.63
cork	4	kurk	5	3.63
death	235	dood	345	3.75
chance	149	kans	171	3.75
rain	74	regen	53	3.75
luck	46	geluk	105	3.75
card	45	kaart	53	3.75
devil	27	duivel	37	3.75
saddle	9	zadel	11	3.75
head	480	hoofd	515	3.88
king	93	koning	87	3.88
tooth	14	tand	12	3.88
breast	46	borst	70	4.00
rhythm	20	ritme	21	4.00
short	196	kort	174	4.13
summer	124	zomer	68	4.13
cool	57	koel	26	4.13
wheel	28	wiel	7	4.13
price	92	prijs	75	4.25
sword	14	zwaard	12	4.25
choir	7	koor	10	4.25
east	182	oost	10	4.38
kiss	29	kus	17	4.38
shoe	15	schoen	10	4.38
breeze	11	bries	4	4.38
cord	8	koord	6	4.38
coffee	92	koffie	111	4.50
pain	78	pijn	149	4.50

Table A1: (Continued)

English Word	English Frequency	Dutch Word	Dutch Frequency	Orthographic Similarity rating
snow	59	sneeuw	39	4.50
crown	23	kroon	23	4.50
bride	11	bruid	10	4.50
deaf	10	doof	8	4.50
jewel	3	juweel	3	4.50
palace	44	paleis	27	4.63
throne	10	troon	11	4.63
pearl	5	parel	3	4.63
year	477	jaar	734	4.75
point	366	punt	144	4.75
gold	92	goud	37	4.75
stone	90	steen	58	4.75
unit	63	eenheid	57	4.75
thief	6	dief	8	4.75
thick	69	dik	44	4.88
moon	55	maan	62	4.88
pure	46	puur	17	4.88
tongue	35	tong	49	4.88
hope	178	hoop	149	5.00
circle	49	cirkel	20	5.00
advice	72	advies	37	5.13
prince	34	prins	59	5.13
grave	32	graf	29	5.13
banana	4	banaan	2	5.13
wound	24	wond	16	5.25
guitar	6	gitaar	5	5.25
total	140	totaal	86	5.38
valley	51	vallei	8	5.38
melon	2	meloen	2	5.38
light	303	licht	339	5.50
street	264	straat	147	5.50
book	254	boek	250	5.50

Table A1: (Continued)

English Word	English Frequency	Dutch Word	Dutch Frequency	Orthographic Similarity rating
green	155	groen	33	5.50
train	77	trein	70	5.50
ship	46	schip	77	5.50
soup	21	soep	22	5.50
domain	12	domein	13	5.50
tomato	7	tomaat	2	5.50
idea	261	idee	150	5.63
leader	69	leider	41	5.63
length	69	lengte	26	5.63
mask	14	masker	14	5.63
hell	97	hel	24	5.75
clock	37	klok	26	5.75
logic	23	logica	17	5.75
baker	16	bakker	12	5.75
idiot	10	idioot	16	5.75
bamboo	6	bamboe	3	5.75
sock	3	sok	2	5.75
doctor	136	dokter	130	5.88
hunger	25	honger	50	5.88
warmth	25	warmte	47	5.88
beard	23	baard	19	5.88
myth	20	mythe	9	5.88
tender	20	teder	12	5.88
fatal	16	fataal	4	5.88
lamb	16	lam	6	5.88
glass	132	glas	124	6.00
milk	109	melk	51	6.00
ball	97	bal	22	6.00
metal	47	metaal	12	6.00
debate	43	debat	11	6.00
pill	14	pil	12	6.00

Table A1: (Continued)

English Word	English Frequency	Dutch Word	Dutch Frequency	Orthographic Similarity rating
mass	111	massa	48	6.13
jury	30	jury	3	6.88
water	452	water	353	7.00
school	368	school	202	7.00
moment	319	moment	270	7.00
hard	271	hard	128	7.00
wind	116	wind	106	7.00
hotel	127	hotel	78	7.00
plan	103	plan	137	7.00
wild	91	wild	38	7.00
type	85	type	53	7.00
winter	83	winter	51	7.00
plant	75	plant	37	7.00
ring	66	ring	24	7.00
fruit	60	fruit	13	7.00
crisis	59	crisis	35	7.00
model	54	model	73	7.00
detail	49	detail	13	7.00
storm	31	storm	27	7.00
sport	31	sport	35	7.00
mild	26	mild	8	7.00
code	25	code	12	7.00
alarm	24	alarm	6	7.00
lamp	23	lamp	21	7.00
drama	22	drama	14	7.00
tennis	22	tennis	2	7.00
oven	19	oven	11	7.00
chaos	16	chaos	18	7.00
circus	15	circus	6	7.00
nest	14	nest	19	7.00
echo	12	echo	8	7.00
menu	8	menu	6	7.00

Table A2: Control stimuli in the language decision experiment

English Word	English Frequency	Dutch Word	Dutch Frequency	Orthographic Similarity rating
soft	82	zacht	114	1.50
army	113	leger	61	1.00
knife	38	mes	33	1.00
cave	29	grot	11	1.13
button	16	knoop	13	N.A.
spark	5	vonk	4	1.38
throat	46	keel	60	1.00
watch	110	horloge	32	1.00
eagle	8	arend	3	1.88
cattle	34	vee	17	1.00
thigh	14	dij	6	1.88
case	376	geval	411	1.00
story	166	verhaal	161	1.00
joke	33	grap	15	1.00
herb	10	kruid	4	1.25
garden	117	tuin	98	1.13
shop	86	winkel	37	1.13
design	81	ontwerp	26	1.13
hole	59	gat	41	1.00
rail	19	spoor	49	1.00
piece	115	stuk	208	1.00
bucket	14	emmer	13	1.25
donkey	10	ezel	8	1.00
pigeon	4	duif	8	1.00
favour	67	gunst	9	1.00
dirt	21	vuil	22	1.00
acid	22	zuur	14	1.00
boot	10	laars	4	1.63
food	266	eten	207	N.A.
crowd	51	menigte	30	1.00
gate	51	poort	27	1.00

Table A2: (Continued)

English Word	English Frequency	Dutch Word	Dutch Frequency	Orthographic Similarity rating
tenant	7	huurder	3	1.00
attack	114	aanval	97	N.A.
engine	44	motor	37	1.00
orphan	3	wees	120	1.00
glue	3	lijm	7	1.38
air	264	lucht	181	1.25
fire	161	vuur	92	1.50
angry	68	boos	41	1.00
duke	39	hertog	23	1.00
limit	37	grens	60	1.25
farmer	33	boer	46	1.13
arrow	8	pijl	9	1.00
free	211	vrij	217	N.A.
fast	104	snel	286	1.13
poem	14	gedicht	31	1.00
painting	66	schilderij	18	N.A.
angle	21	hoek	89	1.00
wife	218	vrouw	597	1.63
animal	120	dier	84	1.00
faith	51	geloof	222	1.00
muscle	33	spier	6	1.00
choice	102	keuze	78	1.75
treaty	16	verdrag	18	1.00
swamp	5	moeras	6	1.00
member	94	lid	114	1.00
song	33	lied	20	1.38
poet	17	dichter	74	1.00
lazy	13	lui	21	1.38
granny	7	oma	19	1.00
peace	92	vrede	51	1.13
game	97	spel	155	N.A.

Table A2: (Continued)

English Word	English Frequency	Dutch Word	Dutch Frequency	Orthographic Similarity rating
branch	56	tak	18	1.13
cow	23	koe	15	N.A.
spoon	12	lepel	11	1.00
duck	8	eend	12	1.00
itch	1	jeuk	5	1.00
virgin	19	maagd	13	1.00
sleeve	10	mouw	14	1.00
monkey	9	aap	12	1.00
small	537	klein	192	1.13
large	373	groot	386	1.00
demand	95	eis	134	1.00
bottle	88	fles	74	1.00
judge	58	rechter	79	1.00
napkin	5	servet	4	1.13
noise	63	lawaai	31	1.13
enemy	53	vijand	40	1.00
guilt	39	schuld	81	1.25
dull	34	saai	8	1.00
road	212	weg	839	1.25
paint	41	verf	26	1.00
shape	66	vorm	242	1.00
target	35	doel	148	1.13
chain	34	keten	8	1.63
carrot	3	wortel	13	1.13
stomach	42	maag	38	N.A.
bullet	14	kogel	16	1.00
doubt	154	twijfel	65	1.13
silly	45	dom	34	1.00
bird	44	vogel	35	1.00
mind	351	geest	175	1.00
body	308	lichaam	264	1.00
office	255	kantoor	56	1.00

Table A2: (Continued)

English Word	English Frequency	Dutch Word	Dutch Frequency	Orthographic Similarity rating
church	163	kerk	170	1.50
danger	76	gevaar	98	1.50
bird	44	vogel	35	1.00
error	21	fout	43	1.25
bull	28	stier	11	1.00
knight	6	ridder	7	1.13
girl	287	meisje	237	1.00
uncle	62	oom	159	1.00
message	71	bericht	38	N.A.
screw	15	schroef	2	1.88
huge	112	enorm	33	1.00
skirt	21	rok	21	1.13
flower	15	bloem	27	N.A.
rifle	17	geweer	32	1.00
alley	10	steeg	28	1.13
pencil	16	potlood	10	N.A.
bright	80	helder	45	1.00
heavy	138	zwaar	92	1.25
proof	32	bewijs	43	1.00
cheese	29	kaas	43	1.88
debt	26	schuld	81	1.25
mirror	43	spiegel	43	N.A.
silk	26	zijde	59	1.00
cage	13	kooi	18	1.63
regret	19	spijt	64	1.13
sure	292	zeker	447	1.13
window	139	raam	112	1.00
sign	106	teken	64	1.13
rent	40	huur	11	1.00
rabbit	11	konijn	10	1.25
pants	17	broek	56	1.00

Table A2: (Continued)

English Word	English Frequency	Dutch Word	Dutch Frequency	Orthographic Similarity rating
cherry	6	kers	1	1.25
wing	33	vleugel	15	1.00
face	472	gezicht	448	1.00
money	390	geld	276	1.00
woman	351	vrouw	597	1.63
power	331	macht	179	1.38
wall	139	muur	91	1.00
cause	127	oorzaak	79	1.00
chair	114	stoel	117	1.00
horse	89	paard	99	1.13
empty	86	leeg	68	1.00
loss	82	verlies	38	1.75
duty	68	plicht	28	1.00
pocket	59	zak	66	1.00
coat	57	jas	42	1.00
vote	55	stem	265	1.00
crime	49	misdaad	16	1.00
blanket	17	deken	20	1.00
fate	35	noodlot	12	1.00
autumn	35	herfst	22	1.00
crazy	33	gek	114	1.00
ease	32	gemak	41	1.00
trace	28	spoor	49	1.25
fever	27	koorts	21	1.00
rumour	10	gerucht	11	1.13
ugly	25	lelijk	24	1.00
voyage	9	reis	82	1.00
tale	17	verhaal	161	1.13
witch	16	heks	11	1.25
mercy	16	genade	21	1.00
pillow	15	kussen	34	1.00
fox	10	vos	5	N.A.