Running head: SECONDARY FAMILY SIZE

Effects of primary and secondary morphological family size in monolingual and bilingual word processing

Kimberley Mulder[a,b], Ton Dijkstra[a], Robert Schreuder[a], & Harald R. Baayen[cd]

[a] Radboud University Nijmegen, Donders Institute for Brain, Cognition, and Behaviour, Montessorilaan 3, 6525 HR, Nijmegen, The Netherlands

[b] International Max Planck Research School for Languages Sciences, Wundtlaan 1, 6525 XD, Nijmegen, The Netherlands

[c] Eberhard Karls University, Wilhelmstrasse 19, 72074, Tübingen, Germany

[d] University of Alberta, Edmonton, Canada

Direct correspondence to:

Kimberley Mulder

Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour

Montessorilaan 3, 6525 HR, Nijmegen, the Netherlands

E-mail: K.Mulder@donders.ru.nl

Phone: +31 24 3617701/ Fax: +31 24 3616066

Abstract

This study investigated primary and secondary morphological family size effects in monolingual and bilingual processing, combining experimentation with computational modelling. Family size effects were investigated in an English lexical decision task for Dutch-English bilinguals and English monolinguals using the same materials. To account for the possibility that family size effects may only show up in words that resemble words in the native language of the bilinguals, the materials included, in addition to purely English items, Dutch-English cognates (identical and non-identical in form). As expected, the monolingual data revealed facilitatory effects of English primary family size. Moreover, while the monolingual data did not show a main effect of cognate status, only form-identical cognates revealed an inhibitory effect of English secondary family size. The bilingual data showed stronger facilitation for identical cognates, but as for monolinguals, this effect was attenuated for words with a large secondary family size. In all, the Dutch-English primary and secondary family size effects in bilinguals were strikingly similar to those of monolinguals. Computational simulations suggest that the primary and secondary family size effects can be understood in terms of discriminative learning of the English lexicon.

Keywords: morphological family size, cognates, bilingual word processing, discriminative learning

Introduction

Reading a word is not just looking up this word in a dictionary. If it were that simple, word processing would be affected only by the number of words that share their beginnings and not by the word's more complex relationships to other words in the lexicon on dimensions such as orthographic or semantic relatedness. It turns out that during reading a word activates not only its own representation in the mental lexicon, but many other lexical representations as well, via a system of relationships that are not necessarily strictly word-form related. Words are not isolated units, but parts of larger networks. In the present study, we focus on the activation of morphological networks in the monolingual and bilingual mental lexicon during visual word processing.

Many behavioural and neurolinguistic studies have investigated the processing consequences of various relationships between words in the mental lexicon, with a great deal of attention directed towards orthographic relations between words (see Andrews, 1997, for an overview of studies on orthographic neighborhood size). Recently, research has also focused on morphological relationships between words in the lexicon. One of these morphological relationships, called ‘morphological family size’, is defined as the number of morphologically related complex words in which a given word occurs as a constituent (Schreuder & Baayen, 1997). For instance, *heartless* and *heartache* are family members of the word *heart*. Words can differ considerably in their productivity in terms of the number of their morphological family members. For instance, the word *house* occurs in more than 30 morphologically related complex words (among which, for example, *house hold*, *garden house*, and *housing*), whereas the morphological family of *horizon* is restricted to only a few words (such as *horizontal).*

Schreuder and Baayen (1997) showed that Dutch words with larger morphological families were processed faster and more accurately in a Dutch visual lexical decision task than Dutch words with smaller morphological families. The facilitatory effect of family size has been replicated for Dutch (Bertram, Baayen, & Schreuder, 2000, De Jong, Schreuder & Baayen, 2000; De Jong, 2002; Kuperman, Schreuder, Bertram, & Baayen, 2009), German (Lüdeling & De Jong, 2002), and English (Baayen, Lieber, and Schreuder, 1997; De Jong, Feldman, Schreuder, Pastizzo, & Baayen, 2002; Juhasz & Berkowitz, 2011). Moreover, several non-Germanic languages also revealed similar effects of family size (see Feldman & Siok, 1997, for Chinese; Moscoso del Prado Martín, Bertram, Häikiö, Schreuder, & Baayen, 2003; Kuperman, Bertram, & Baayen, 2008, for Finnish; Moscoso del Prado Martín, Deutsch, Frost, de Jong, Schreuder, & Baayen, 2005, for Hebrew; Boudelaa & Marslen-Wilson, 2011, for Arabic). Importantly, the family size effect is observed to be predictive over and above other lexical properties such as word frequency, morpheme frequency, word length, orthographic neighborhood size, bigram frequency (De Jong et al., 2000, Schreuder & Baayen, 1997), and age of acquisition (De Jong, 2002).

The traditional interpretation of the morphological family size effect holds that upon reading a word, many of its morphological family members become activated thanks to shared orthography, morphology, and semantics (Schreuder & Baayen, 1997). More specifically, activation is assumed to spread from a target word to its family members via direct semantic and orthographic connections. Schreuder and Baayen (1997) proposed to understand the family size effect along the lines of the multiple read-out model of Grainger and Jacobs (1996): Words that co-activate many other words (lemmas) give rise to more global lexical activation supporting a positive lexicality decision. By means of a computational simulation study, De Jong, Schreuder, and Baayen (2003) showed that read-out of global activation may not be necessary if activation is allowed to resonate between forms, lemmas, and meanings.

An unresolved question is whether activation can spread beyond immediately related concepts to concepts that are only indirectly linked to a target word. Studies of mediated priming have demonstrated that a target word such as *cheese* can be processed faster when it is preceded by a prime such as *cat* that is only indirectly related to the target in semantic memory via a mediating concept (*mouse*) than when it is preceded by a semantically unrelated prime (e.g., *table;* cf. De Groot, 1983). Mediated priming effects were observed in word naming (Balota & Lorch, 1986), in a double lexical decision task in which a lexical decision to both the target and prime is required and in which only indirectly related prime-target pairs were used, and in a single presentation lexical decision task in which the prime and target were presented with no obvious pairing and a lexical decision was required to both items (McNamara and Altarriba, 1988). However, a number of studies failed to find the mediated priming effect in standard lexical decision (e.g., Balota & Lorch, 1986; Chwilla, Kolk, & Mulder, 2000). As Chwilla et al. (2000) argued, mediated priming seems to occur only when the lexicality of both the prime and the target needs to be judged. In sum, these studies show that activation can spread beyond directly related concepts, albeit only under special experimental conditions. Applying this idea of spreading of activation to the case of family size, it is conceivable that activation spreads from immediate family members, which are directly related to the target in form and meaning, to more distant family members at greater distances in the lexical network, i.e., to words that are related to the target word only via their primary family members.

Recent studies (Baayen, 2010a, and Baayen, Milin, Filipovic-Durdevic, Hendrix, & Marelli, 2011) indicate that more distant morphological relatives can influence compound processing. These studies propose a new measure, the secondary family size, as a means for gauging the relevance of more distant morphological relatives. Recall that the primary family size of a given noun contains all words, both derived words and compounds (except the noun itself) that contain that noun as a constituent. Baayen (2010b) and Baayen et al. (2011) argued that although the primary family size is defined across both derived words and compound words, most of a given word's family members are compounds. In these studies, the secondary family measure was therefore operationalized on the set of compounds, and was further restricted to family members that are two-constituent compounds. In the present study, the focus is on the processing of monomorphemic words, and hence, a definition of secondary family including both compounds and derivations is applied. Informally, the secondary family size of a word can be defined as including all words that share a constituent with a word in a word's primary family, excluding the primary family members themselves (for a formal definition of secondary family size, see the Appendix). Figure 1 presents a schematic representation of the activation of primary and secondary family members of the target word *horse*.

(Figure 1 about here)

If activation spreads from a target word, first into the primary family, and then on into the secondary family, the question arises whether the co-activation of secondary family members is facilitatory (just like the primary family size) or rather inhibitory. Theories restricting primary and secondary family size effects to the level of word form offer no prediction. Because activating primary family member word forms is facilitatory in lexical decision, activating even more word forms might also speed up ‘yes’ responses in this task. Alternatively, it is conceivable that activating many orthographically unrelated word forms (such as *hairbrush* for *horse*) would, due to feedback connections, reduce the bottom-up support from the letter layer to the word layer for the target word. For instance, the *h* and *r* in *horse* might become, due to spreading activation, more ambiguous between *horse* and *hairbrush*, and would therefore delay lexicality decisions.

However, theories seeking to explain the primary and secondary family size effects at the level of semantics make a clear prediction. The primary family members are semantically related to the target. Knowing what a *horse* is entails, for instance, knowing that horses have to deal with *horseflies*. The secondary family members tend not to be semantically related. A *workbox* is a box storing tools for sewing, a *cocktail* is a drink, and horses do not wear *hairnets*. The activation of unrelated meanings should therefore have a detrimental effect on response speed.

For response times to compounds in visual lexical decision as available in the English Lexicon Project (Balota et al., 2007), Baayen (2010b) observed an effect of secondary family size, which was modulated by the size of the primary family of the compound's head, and by the density of the compound graph (operationalized by the graph-theoretical concept of the strongly connected component, i.e., the subgraph for which it holds that any constituent can be reached by following the directed edges connecting modifiers to heads). The predicted inhibitory effect of secondary family size was present for compounds with a smaller right constituent family size, and most strongly so for compounds that were not part of the strongly connected component of the compound graph.

An inhibitory effect of secondary family size fits well within a semantic explanation of the family size effect. There is increasing evidence that the family size effect is at least partially semantic in nature. Schreuder and Baayen (1997) observed that positive correlations between family size and reaction times increased when semantically opaque family members were excluded from the family size count (e.g., *honeymoon* is morphologically but not semantically related to *honey*; exclusion of opaque family members such as *honeymoon* from the family size count of *honey* increased the positive correlation of family size with RT).

Moreover, De Jong et al. (2000) showed that the family size effect appeared for both regular and irregular past participles (e.g., *roei-geroeid*, ‘row-rowed’ vs. *v**e**cht-gev**o**chten*, ‘fight-fought’, even though the irregular past participle does not share the exact form with its mono-morphemic stem and other family members. Again, inclusion of a morphologically related but not semantically related form such as *vocht* (meaning ‘moisture’) in the family size count of *vecht* decreased the correlation between RTs and family size.

Moscoso del Prado Martín et al. (2005) report an additional semantic characteristic of the family size effect in Hebrew. They observed that activated semantic fields of morphological roots that were related in meaning to a Hebrew word had a different effect on response latencies than unrelated activated semantic fields. In a Hebrew visual lexical decision task, Moscoso et al. not only observed the expected facilitation effect of family members that were related in meaning, but they also observed an inhibition of RTs when the number of family members that were not semantically related increased.

Finally, in an ERP study with Dutch monolinguals, Mulder, Schreuder, and Dijkstra (2012, Experiment 2) observed less negative N400 amplitudes for Dutch words with a large Dutch primary family size compared to words with a small Dutch primary family size. They pointed out that the observed pattern for activated family members is different from the ERP effects reported in the literature for orthographic neighbors and semantic associates (Müller, Duñabeitia, & Carreiras, 2010), because the latter activate semantic representations that are different or less compatible with that of the target, while primary family members always activate compatible semantic representations.

In sum, these studies show that the family size effect is at least partially semantic in nature. Moreover, the different effects for semantically related and unrelated family members observed by Moscoso et al. (2005) and Mulder et al. (2012) give rise to the hypothesis that semantic overlap between target word and family member can determine the direction of the family size effect. Apparently, if activation spreads too far out and reaches semantically unrelated words, then facilitation reverses into inhibition.

Until now, not many studies have investigated family size effects in bilinguals. During the acquisition of a second language (L2), bilinguals will learn new words and consequently start to develop morphological and semantic relationships between those words in their L2. It is therefore likely that the primary family size of the L2 starts affecting L2 word processing, even though the primary family size of words of their L2 may be not as large in the lexicon of bilinguals as the primary family size of words of their first language (L1). Moreover, if lexical activation spreads to more distant family members, as is observed in monolingual processing by Baayen (2010b), even L2 secondary family members should be activated and affect L2 word processing.

Dijkstra, Moscoso del Prado Martín, Schulpen, Schreuder, and Baayen (2005) investigated the role of L1 and L2 primary family size in the processing of Dutch-English interlingual homographs (e.g., *room*, meaning 'cream' in Dutch) by Dutch-English bilinguals. First, they conducted a re-analysis of available English (L2) lexical decision data from Dutch-English bilinguals by Schulpen, Dijkstra, and Schriefers (2003), which included both purely English words and Dutch-English interlingual homographs. This re-analysis revealed a facilitatory effect of L2 family size on the processing of purely English words and Dutch-English interlingual homographs. Furthermore, the interlingual homographs also showed inhibitory effects of the family size of the non-target language, Dutch (L1). The observed morphological family size effects were independent of the relative frequency of the two readings of the homographs. Interestingly, the same pattern was found when bilinguals made lexical decisions on interlingual homographs in their L1: Facilitation of the target language (Dutch) and inhibition of the non-target language (English). This study shows that bilinguals are sensitive to the primary morphological productivity of words of both the target and non-target language when reading in only one language. Moreover, the findings that activation of the non-target language family members of Dutch-English interlingual homographs in language-specific lexical decision inhibits target word processing supports the hypothesis that family size effects are mediated by semantic similarity.

Further bilingual evidence comes from Mulder et al. (2012, Experiments 3 and 4), who observed that Dutch-English bilinguals activate the cross-language (English) primary family size for Dutch-English cognates in a Dutch task context. Similar to the pattern of within-language effects observed for Dutch monolinguals (Experiments 1 and 2), a large cross-language family size led to faster response latencies in Dutch lexical decision task and less negative N400 amplitudes in a Dutch go/no-go task while ERPs were recorded. Also, the ERP effects for cross-language family size were different from effects for cross-language neighborhood size observed in the literature and support the semantic interpretation of the family size effect that was outlined above.

The aim of the present study is to investigate whether and how extensively, during L2 word processing, activation spreads within the bilingual mental lexicon. More specifically, we want to investigate whether the secondary family size of L2 items affects L2 word processing or whether it is only the L2 primary family size that is activated. The literature on mediated priming and the secondary family size effects in the monolingual data reported by Baayen (2010b) suggest that even distantly related lexical items can become activated during word processing in isolation. Moreover, the bilingual data of Dijkstra et al. (2005) show that bilinguals are sensitive to the primary morphological productivity of L2 items. However, assuming that the links between English words are less strong for Dutch-English bilinguals compared to English monolinguals, it is not evident that lexical activation in their second language spreads beyond directly related items.

Effects of secondary family size may even only affect items that have a strong representation in the bilingual lexicon, such as cognates. Cognates are words in both languages of a bilingual that share most of their form and meaning in these languages. Just because of their 'double nationality', cognates may be more strongly represented, and more easily accessed than words of similar frequency that belong to one language only. Cognates can be either identical in form (e.g. *hotel* in English and Dutch) or nearly identical (e.g. *altar-altaar* in English and Dutch, respectively). Bilingual research has shown that reading a cognate co-activates the target language and non-target language lexical representations of the cognate (see Dijkstra, 2005, for an overview of studies). In line with this observation, it has been proposed (Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010) that cognates are characterized by two overlapping orthographic representations that are linked to a (largely) shared semantic representation. The observation of a cognate effect (i.e., faster RTs to cognates than to non-cognates) can then be explained by a combination of co-activation and orthographic-semantic resonance. Reading a cognate will lead to co-activation of two overlapping orthographic representations, which will activate their corresponding primary and secondary family members in both languages. Reading a non-cognate, however, activates only one representation and its primary and secondary morphological family in only one language. As a consequence of the co-activation in cognates, which will activate a (largely) shared semantic representation, activation can pass more easily to other, more distant, items of the target language during word processing, strengthening the activation of the target language secondary family. Thus, activation of target language secondary family members is more likely to be observed for cognates than for non-cognates. In addition, most co-activation is expected for cognates that have complete form overlap with words in their first language (i.e., identical cognates). Therefore, in this study, the stimulus materials will include both identical and non-identical cognates, in addition to purely English words.

In Experiment 1, we sought to replicate the effects of primary and secondary family size observed in monolingual research with our set of cognate and non-cognate items. Replicating the secondary family size effects reported by Baayen (2010b) is of particular interest here, because, to date, these effects have not been replicated with new empirical data. This was accomplished by means of an English visual lexical decision task with English monolinguals. We expected that the distinction between cognates and non-cognates would be irrelevant for monolinguals, and therefore expected family size effects to affect the processing of cognates and non-cognates in the same way. In Experiment 2, the same task with the same materials was performed by Dutch-English bilinguals. To our knowledge, this is the first study that directly compares both primary and secondary family size effects in monolingual and bilingual processing. Moreover, this is the first study that addresses L2 family size effects in cognates.

After having reported the experimental results, we compare two theoretical frameworks for understanding the primary and secondary family size effects: the general framework of spreading activation and the more recently developed framework of discrimination learning. Over the years, spreading activation has proven to be a fruitful paradigm to investigate word processing, with influential interactive activation models such as IA and BIA (McClelland & Rumelhart, 1981; Dijkstra & Van Heuven, 1998), and the multiple read-out model (Grainger & Jacobs, 1996) being able to account for a wide range of effects. However, the non-interactive framework of naïve discrimination learning (Baayen et al., 2011) provides an alternative account of many previous findings on morphological processing. By means of computational simulation studies of the data of Experiments 1 and 2 with naïve discrimination learning, we will examine whether this type of approach is as successful as interactive activation models in explaining the present experimental data.

Before we turn to the two experiments and the modeling section, we will first discuss how family size measures were improved for use in our experiments.

Family Size Generation Study

A major resource for researchers working on morphological family size is the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). CELEX provides family size counts for English, Dutch, and German. These counts are highly informative and have proven to be useful in past and present-day research on family size. However, the CELEX database does not provide realistic frequency information for English spaced compounds (all have a frequency of zero) and therefore these are not included into the family size count of this database. Therefore, these counts may not provide a realistic representation of family size counts for speakers of English.

To improve the existing CELEX primary family size counts, we let Dutch (L1)-English (L2) bilinguals perform a Family Size Generation task in which they had to produce morphological family members for a list of English target words. These data were used to create a primary family size measure based on both the original CELEX count and the count obtained by the Family Size Generation task. We chose to select Dutch-English bilinguals for the generation of the family members, because the focus of the study is on family size effects in bilingual word processing. Inclusion of the most frequently generated spaced compounds known by bilinguals in the English family size count will likely result in a more accurate family size count for this participant group and a better prediction of response latencies in bilingual word processing. Moreover, we expect that this measure improves the available family size counts as provided by CELEX even for monolingual word processing, because (the most frequent) spaced compounds are now added to the existing count. When English monolinguals generate family members for English, this will probably result in family size counts that are overestimated relative to Dutch-English bilinguals (see argumentation in General Discussion). Nevertheless, for the purpose of checking and comparison, we also asked a group of English monolinguals to generate morphological family members for the set of cognates in their native language.

Method

Participants. Forty-five Dutch L2 speakers of English (mean age= 22.6, SD = 3.49), mostly undergraduates at the University of Nijmegen, were paid to take part in this Generation Study. All were highly proficient in English, having learned English at school from the age of 11. All participants had normal or corrected-to-normal vision. Twenty-one L1 speakers of English (mean age = 21.4 , SD = 3.69), were recruited at the University of Nottingham. None of the participants had any knowledge of Dutch. The participants reported not to have any substantial active knowledge of other languages, although it is likely that they have received some education in another foreign language at secondary school.

Materials. For the Generation Study, all word items that were to be used as experimental items in Experiments 1 and 2 were selected. A list of the items is provided in the Appendix. All items were monomorphemic nouns that did not have a homographic conversion verb. The length of the items ranged between three and eight letters.

We divided the stimuli over three lists. To obtain an equal number of stimuli in all lists and to be able to compare the lists in each version, we added some filler items that were the same in each version. The total number of items in the English lists was 50. The items of the three English lists were matched on English log lemma frequency per million and log English CELEX family size as much as possible. For the English monolingual participants, there was only one stimulus list containing the 50 cognate items.

Procedure. The Dutch-English bilingual participants were tested in a noise-proof experimental room. They saw only words of one of the lists. The lists were randomized for all participants. Participants were given a list of stimulus words and were asked, for each stimulus word on the list, to generate other words in which the stimulus word could occur. The items were presented in capital letters in an Excel file on a HP Compaq Intel Core 2 computer. Participants were asked to type the words in the fields directly following the target word. It was emphasized that they could write down a word even if they were not confident of the exact orthography of that word. Furthermore, they were told that they were allowed to skip a target word when they could not think of any words for that target word and return to that target word when they came up with new words. A time limit of thirty minutes was set to complete the task. A pilot experiment showed that this amount of time was enough for participants to respond to all the items and go through the list again to see if they could come up with some more words.

The procedure for the English monolingual participants was identical to that of the bilingual participants, except that the participants wrote down the family members on a piece of paper instead of in an excel sheet.

Results

For each item, we listed all family members that were generated. We did not consider inflected words (e.g., *houses* is not counted as a family member of *house*), and only included compounds and derivations (e.g., both *normal* and *age norm* are family members of *norm*). Finally, for each target item we counted the number of different words that were generated.

We then selected for each item those family members that were generated by at least three participants in order to include in our family size count only well-known family members and to exclude very low frequent family members. Next, we checked whether these family members were present in the CELEX count, and if this was not the case, we added these items to the CELEX count. In this way, an “updated” version of the CELEX count was obtained containing family members that are nowadays commonly used but that were missing in the CELEX count (see the Appendix for the new family size values). The correlation between the CELEX English Family Size counts and the new English family size measure based on the bilingual counts (from now on, *English Primary Family Size)* was .87. Furthermore, the correlation of CELEX English Family Size with the mean lexical decision latencies from the English lexicon project (Balota et al., 2007) was -0.20. When replaced by our new measure, *English Primary Family Size*, this correlation increased to -0.29. The family size count for the cognates generated by the English monolinguals correlated well with the count obtained from the bilinguals (r = .91).

Discussion

The purpose of the Family Size Generation task was to improve the existing English primary family size count as provided by the CELEX lexical database. CELEX does not include spaced compounds into the English family size count. Our new family size measure, which includes the most common spaced compounds, is, as we shall see, a better motivated predictor of response latencies than the original CELEX family size measure. Moreover, the high correlation between the bilingual and monolingual counts for the cognate items shows that the new measure can be used with confidence to assess family size effects in both bilingual and monolingual word processing.

In Experiments 1 and 2, we applied the new English family size measure to assess family size effects in monolingual and bilingual language processing. In Experiment 1, we conducted an English lexical decision task with English monolingual speakers. The aim of this experiment was to replicate earlier monolingual research on morphological family size effects in visual word processing reporting facilitation effects of primary family size and inhibitory effects of secondary family size. Replicating the secondary family size effects reported by Baayen (2010b) is of particular interest here, because, to date, these kinds of effects have not been replicated with new empirical data. In this experiment, we included both English-Dutch cognate and non-cognate items. Because the monolingual English speakers should be insensitive to the cognate status of the English items, we predicted no significant effect of cognate status and no interaction of cognate status with either primary or secondary family size.

Experiment 1 – English visual lexical decision with English monolinguals

Method

Participants. Twenty-eight native English speakers (mean age = 21.8 years old, SD = 3.53) were recruited at the University of Nottingham. None of the participants had any knowledge of Dutch. Although it is likely that they have received some education in another foreign language at secondary school, the participants reported not to have any active knowledge of other languages. All participants had normal or corrected-to-normal vision. They were paid or received course credits for their participation.

Materials. The stimulus set consisted of 300 items, half of which were English words and half were non-words. All word items were selected from the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995). Only word items with an English lemma frequency of at least one per million in the CELEX database and a length between three and eight characters were selected. All items were mono-morphemic nouns that had no conversion verb. For each item, the English primary family size values were calculated and logarithmically transformed. The primary family size values were based on the new family size measure (*English Primary Family Size*, see Family Size Generation Study). These family size values were collinear with the values of the logarithmically transformed values of *SBTLWF* (English Subtitle Frequency per million; Brysbaert & New, 2009). Recent research shows that *SBTLWF* is a better predictor of response latencies than the English CELEX frequency measure (Brysbaert & New, 2009). In the remainder of this paper, we will use the term *English Frequency* to refer to the logarithmic transformation of *SBTLWF*. To remove collinearity, we regressed *English Primary Family Size* on *English Frequency* and used the resulting residuals as new predictors of English family size uncontaminated by English frequency.

Secondary Family Size was operationalized on the set of bimorphemic words, including both derivations and compounds. In this respect, we slightly differ from Baayen (2010b) and Baayen et al. (2011) whose family size definition was optimized for the processing of compounds, and therefore only included two-constituent compounds in the secondary family size count. Because the targets in our study are all monomorphemic words, a definition of secondary family size including all morphologically related words, thus including derivations, seems more appropriate. Moreover, in this way, the definitions of primary and secondary family size are more similar. The values for secondary family size were logarithmically transformed. The correlation between the measure of *English Primary Family Size* (residualized on *English Frequency*) and the measure of *Secondary Family Size* is positive, as expected, but with a correlation of $r = 0.47$ ($p < 0.0001$). That is small enough not to require further orthogonalization from the measure of *Primary Family Size*.

The experimental word items were 50 English-Dutch cognates, i.e., translation equivalents that overlap in form. Half of the experimental items were identical cognates (i.e., items that have complete orthographic overlap in English and Dutch, such as *hotel* and *norm*), whereas the other half were non-identical cognates in English and Dutch (e.g., *thief*-*dief* and *planet*-*planeet*). The latter items also shared their orthographic form in both languages, but the overlap was not completely identical and differed on maximally three letter positions. The degree of orthographical overlap was calculated by the Levenshtein distance measure (Levenshtein, 1966). The Levenshtein distance is the minimal number of deletions, insertions, or substitutions that is required to transform the source string into the target string. All cognates were pure noun cognates in the sense that both the English and Dutch word forms only belonged to the class of nouns. The Dutch noun frequency per million was taken from the CELEX database and was logarithmically transformed. It was made sure that these items had a Dutch noun frequency of at least one per million.

For each cognate item, the Dutch frequency and family size values were calculated. The Dutch lemma frequencies per million were extracted from the CELEX database (*Dutch Frequency*). The Dutch family size values (*Dutch Family Size*) were based on type counts of the family members listed in CELEX. Both the frequency and family size values were logarithmically transformed. The Dutch family size values were collinear with the Dutch frequency values. To remove this collinearity, we regressed the family size values on these frequency values and used the resulting residuals as a new predictor of Dutch family size uncontaminated by Dutch frequency. The Dutch secondary family size counts of the items (*Dutch Secondary Family Size*) were obtained by summing the positional family sizes of their family members. The secondary family size values were logarithmically transformed.

The cognate items were matched to 50 control items on *English Primary Family Size*, *English Frequency*, and length in letters. Moreover, the total set of cognate items was matched to the set of control items with respect to *Age of Acquisition* (*AoA*; extracted from Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012; *AoA* ratings available for 99 of the 100 stimuli), and *English Bigram Frequency* (extracted from the database of the English Lexicon Project). Table 1 displays the characteristics of the cognates and controls. The experiment also included 50 filler words and 150 pseudo-words that were matched to the experimental stimuli on length, and for the filler word items also on *English Frequency*. The 150 non-words resembled English words with respect to their orthography and phonology, and were created by replacing one or more letters of existing English words. The experiment consisted of two item blocks. The presentation order of the items within a block was randomized individually and had the restriction that no more than three words or non-words could follow each other directly.

(Table 1 about here)

Procedure. Participants performed an English visual lexical decision task. In this task, participants decide whether or not the visually presented stimulus is an existing English word by pressing a button corresponding to either the answer 'yes' or 'no'. The task was developed and carried out in *Presentation* version 13.0 (Neurobehavioural Systems, www.nbs.com) and was run on a HP Compaq Intel Core 2 computer with 1.58 GHz processing speed and a refresh rate of 120 Hertz. The participants were seated at a table at a 60 cm distance from the computer screen. The visual stimuli were presented in white capital letters (24 points) in font Arial in the middle of the screen on a dark grey background. Participants were tested individually in a soundproof room.

Participants first read the English instructions, which informed them that they would be presented with word strings and which asked them to push the 'yes' button if the letter string they saw was an existing English word and to push the 'no' button if it was not. They were asked to react as accurately and quickly as possible.

Each trial started with the presentation of a black fixation point '+', which was displayed in the middle of the screen for 700 ms. After 300 ms the target stimulus was presented. It remained on the screen until the participant responded or until the timeout at 1500 ms. The visual target stimulus disappeared when the participant pressed the button, or when the time limit of 1500 ms was reached, and a new trial was started after an empty black screen of 500 ms.

The experiment was divided in two parts of equal length. The first part was preceded by 20 practice trials. After the practice trials, the participant could ask questions before continuing with the experimental trials. The two parts each contained 150 experimental trials. Each part began with three dummy trials to avoid lack of attention during the beginning of the two parts. The end of the first part was indicated by a pause screen. The experiment lasted for approximately 16 minutes.

Results

Data cleaning was first carried out based on the error rate for participants and word items. All participants had an error rate of 10% or less on the word items. Therefore, no participant data were removed. The overall error rate on the experimental word items was 3.8% of the total of 2800 data points. Six word items that elicited errors in more than 15% of the trials were removed from the data set. Interestingly, these word items were all cognate words (*chaos*, *norm*, *flora*, *psalm*, *villa,* and *cigar*). RTs from incorrect responses or null responses were removed from the remaining data set (2.39% of the data points). This resulted in a data set with 2569 data points. Inspection of the distribution of the response latencies revealed non-normality. A comparison of a log transform and an inverse transform ($RT_{inverse} = -1000/RT$) revealed that the inverse transform was most successful in reducing this non-normality.

Response latencies were analyzed with a linear mixed effects model with subject and item as crossed random effect factors (see, e.g., Baayen, 2008; Baayen, Davidson, & Bates, 2008). We first fitted a simple main effects model to the data including all 2569 data points.

Besides *English Frequency, English Primary Family Size* and *English Secondary Family Size*, the following other predictors were considered that might affect lexical decision latencies. To assess the value of our new measure of primary family size in comparison to the original CELEX measure, we included the predictor *CELEX Primary Family Size*. Further, in order to test whether cognate items were processed differently from non-cognate items, we included a factor *Cognate* with the levels 'cognate' and 'non-cognate'. Moreover, to account for possible differences between identical cognates and the other stimuli that do not have complete overlap between English and Dutch, the factor *Identical Cognate* (with the levels Identical cognates and Other items (the latter including non-identical cognates and non-cognate controls)) was considered. As further bilingual factors, *Dutch Primary Family Size* and *Dutch Secondary Family Size* were included in the analyses to clarify whether the family size of another language could affect response latencies in English lexical decision. This should obviously not be the case for English monolinguals that have no knowledge of Dutch, but they could affect the responses of Dutch-English bilinguals. Inclusion of these factors increases similarity between the monolingual and bilingual analyses. Furthermore, to be able to remove any auto-correlation from the error, we included *PreviousRT* (the logarithmically transformed response latency at the previous trial) and *Trial* (the rank of the item in the experimental list) as predictors (cf. Baayen, 2008 and Baayen & Milin, 2010). *OLD* (OLD-20; defined as the mean of the closest 20 Levenshtein Distance orthographic neighbors; see Balota et al., 2007, and Yarkoni, Balota, & Yap, 2008) was included as a predictor to account for effects of similarity between English words. Finally, other variables we considered were the number of syllables of the target word (*NSyllables*), whether the initial syllable of the target word was stressed or not (*InitialStress*), the number of English neighbors (*OrthoN*), and the Levenshtein distance between the English and Dutch reading of the word (*Levenshtein*).

We performed a stepwise variable selection procedure in which non-significant predictors were removed to obtain the most parsimonious model. Important to note here is that the predictor *CELEX Primary Family Size* was not significant and did not correlate significantly with the mean lexical decision latencies. When replaced by our new measure, *English Primary Family Size,* there was a significant correlation ($r = 0.20$). Next, potentially harmful outliers (defined as data points with standardized residuals exceeding 2.5 standard deviation units) were removed from the data set. We then fitted a new model with the same significant predictors to this trimmed data set.

The final model incorporated three parameters for the random-effect structure: a standard deviation for the random intercept for item (SD = .09), a standard deviation for the random intercept for subject (SD = .20), and a standard deviation for the by-subject random slopes for *Trial* (SD = .05). Justification for the use of these random-effect factors was provided by likelihood ratio tests (all p-values < .05). Other random-effect parameters were tested, but were not significant. The standard deviation for the residual error was .31. Three predictors (*English Primary Family Size*, *English Frequency,* and *PreviousRT)* reached significance as main effects. In addition, an interaction between *Identical Cognate* (identical cognates versus non-identical cognates and controls) and *English Secondary Family Size* was present. Table 2 summarizes the coefficient of the fixed effects for the resulting model, together with their standard error, t-values, and p-values based on 10,000 MCMC samples from the posterior distribution of the parameters. Figure 2 visualizes the significant partial effects of *English Frequency* (panel a), *PreviousRT* (panel b), and *English Primary Family Size* (panel c) and the interaction of *Identical Cognate* and *English Secondary Family Size*.

(Table 2 and Figure 2 about here)

Both *English Frequency* and *English Primary Family Size* had a facilitatory effect on response latencies. The main effect of *English Secondary Family Size* did not reach significance, but *English Secondary Family* size did emerge in a significant interaction with *Identical Cognate*. Furthermore, *PreviousRT* had a negative correlation with response latencies, showing that a slow response is often preceded by a fast response.

Discussion

In this experiment, we replicated the primary family size effect as reported in earlier monolingual research (e.g. Schreuder & Baayen, 1997; Baayen et al., 1997; Bertram et al., 2000; De Jong, 2002): *English Primary Family Size* had the expected facilitatory effect on response latencies. The primary family size measure based on the counts obtained from the Family Size Generation Study turned out to be a better predictor than the original CELEX measure for primary family size. Whereas the latter predictor was non-significant, our new family size measure did emerge as a significant predictor in the model. This shows that the addition of spaced compounds to the original family size count resulted in an improved predictor of family size effects in monolingual lexical decision.

Both *Dutch Primary Family Size* and *Dutch Secondary Family Size* did not produce significant effects. This is not surprising given the fact that the English monolinguals in our study did not have any knowledge of Dutch. Hence, they should neither process cognates differently from controls, nor should they be sensitive to the morphological productivity in Dutch of the cognate items.

There was no main effect of *English Secondary Family Size*, but this variable turned out to interact significantly with a variable distinguishing between identical cognates and other stimuli (*Identical Cognate*), showing inhibitory effects in identical cognates but not in the other stimuli. The observed direction of the effect is in line with Baayen (2010b), who also observed that large secondary family sizes can slow lexical processing. Because most of a word's secondary family members are not semantically related to its meaning, activation of these secondary family members will interfere with the interpretation of the presented stimulus.

An effect of secondary family size that emerges only for the identical cognates was not predicted. This finding challenges the assumption of simple spreading of activation, because in this view activation is expected to spread to all items, to both cognates and controls. As the English monolinguals are insensitive to the cognate status of the items, an explanation of the interaction would logically not involve language membership of the items but should be sought elsewhere. Because the identical cognates, the non-identical cognates, and the controls were carefully matched for primary and secondary family size, frequency, length, and bigram frequency (see Table 1), we can rule out that an imbalance in, say, primary productivity would be at issue. Morever, it was suggested by one of our Reviewers that a possible imbalance in *Age of Acquisition* might explain why the interaction of secondary family size with identical cognate is observed in monolinguals. Identical cognates indeed slightly differ from the total set of non-identical cognates and controls ($p = .045$, but they were matched on *AoA* to non-identical cognates ($p > .05$)) and tend to be acquired at a somewhat later age. However, this is in contradiction with the faster RTs for identical cognates, and conflicts with an explanation in terms of *AoA*. Also, effects of *AoA* are largely explained by other variables that we did include in our model, such as frequency (cf. Baayen, Feldman, & Schreuder, 2006). Moreover, as we shall see later in this paper, in a joint analysis of the monolingual and bilingual data, and in a distributional analysis of RT data from the English Lexicon Project, there is no confound between *AoA* and cognate status nor can *AoA* predict the cognate status of a word. We therefore believe that *AoA* cannot offer an explanation for the observed interaction between *Identical Cognate* and *Secondary Family Size*. However, an additional role of yet other variables (e.g., imagery) cannot be excluded and should be topic of future investigation.

Importantly, the interaction of secondary family size with identical cognates does not logically entail that the monolinguals were sensitive to the historical origin of the identical cognates, but rather that these subjects were sensitive to the specific distributional characteristics of the mapping of form characteristics to meanings. Anticipating the results of our computational modeling to be discussed below, it turns out that this interaction falls out as a straightforward consequence of the distributional properties of English. First, however, we consider whether Dutch-English bilinguals show the same pattern of results for this set of stimuli: facilitation from the primary family size, but inhibition from the secondary family size for identical cognates only.

In Experiment 2, we used the same materials in an English lexical decision task, this time with Dutch-English bilinguals. Having developed morphological and semantic relationships between words from their L2, English, these bilinguals should activate morphological family members of English words. Although the morphological family size of English words might be lower for bilingual than for monolingual speakers, English primary family size is expected to affect bilingual word processing in a way similar to monolingual processing, facilitating comprehension. Moreover, if the participants are sufficiently proficient, secondary family size effects might also be visible, in which case it should be restricted to the identical cognates only.

In addition, assuming that the bilinguals activate both target and non-target representations when reading a cognate, we will consider non-target language (Dutch) primary and secondary family size effects in the set of cognates as well. Given the semantic overlap between the Dutch family members and the cognate target word, we expect that the direction of the Dutch primary and secondary family size effect patterns with the effect of English primary and secondary family size.

Experiment 2 – English visual lexical decision with Dutch-English bilinguals

Method

Participants. Thirty-three students of the University of Nijmegen (mean age 22.8 years, SD = 3.48) took part in this experiment. All participants had normal or corrected–to-normal vision and were native speakers of Dutch, having English as their second language. They had learned English at school from around the age of 11. Participants were paid or received course credits for participating in the experiment.

Materials. The 50 cognate and 50 non-cognate control items were identical to those used in Experiment 1. The experiment further included 50 English filler words and 150 pseudo words that were matched to the experimental stimuli in length, and for the filler word items, also in English frequency.

Procedure. The procedure of the lexical decision task is identical to the procedure of Experiment 1. After completing the lexical decision task, participants performed the LexTALE task (Lemhöfer & Broersma, 2012). This task was used to obtain a general indication of their proficiency in English in terms of vocabulary knowledge. Finally, participants were asked to fill out a language background questionnaire. The total session lasted approximately 25 minutes.

Results

Data cleaning was first carried out based on the error rate for participants and word items. Participants with an error rate of more than 15% on the word items were removed from the data set, which resulted in the exclusion of the data from three participants.

Eleven word items (cognates: *baron, flora, norm, cigar, pill,* controls: *dusk, cattle, thigh, cellar, lad,* and *torch*) that elicited errors in more than 15% of the trials were removed from the data set. After removal of these items, we were left with 2670 data points on the word items. RTs from incorrect responses or null responses were removed from the remaining data set (2.92% of the data points). This resulted in a data set with 2591 data points. Inspection of the distribution of the response latencies revealed non-normality. A comparison of a log transform and an inverse transform ($RT_{inverse} = -1000/RT$) revealed that the inverse transform was most successful in solving this non-normality.

As before, response latencies were analyzed with a linear mixed effects model with subject and item as crossed random effects. We considered the same predictors as in Experiment 1. Because bilinguals are expected to be sensitive to non-target language frequency and non-target language family size effects, *Dutch Frequency*, *Dutch Primary Family Size,* and *Dutch Secondary Family Size* we also considered as predictors.

To obtain the simplest best fitting model, we applied the same procedure of variable selection and exclusion as in Experiment 1. Potentially harmful outliers (defined as data points with standardized residuals exceeding 2.5 standard deviation units) were removed from the data set. A new model with the same predictors was fit to this trimmed data set. The final model incorporated five parameters for the random-effects structure of the data: a standard deviation for the random intercepts for subject (SD = .18) and item (SD = .08), as well as a standard deviation for the by-subject random slopes for *Identical Cognate* (SD = .07) and *Trial* (SD = .03), and a correlation parameter for the by-subject slope for *Identical Cognate* and the by-subject random intercept (r = .30). The standard deviation for the residual error was .26.

The final model contained five numerical predictors (*English Primary Family Size*, *English Frequency*, *OLD*, *English Secondary Family Size* and *PreviousRT*), one factorial predictor (*Identical Cognate*) and one interaction (*Identical Cognate:English Secondary Family Size*). The relevant statistics and corresponding coefficients of the final model are reported in Table 3. The partial effects of *English Frequency* (panel a), *English Primary Family Size* (panel b), *Identical Cognate* (panel c), *English Secondary Family Size* by *Identical Cognate* (panel d), *OLD* (panel e) and *PreviousRT* (panel f) of the final model are visualized in Figure 3.

(Table 3 and Figure 3 about here)

As expected, we observed facilitatory effects on response latencies for both *English Frequency* and *English Primary Family Size*. Moreover, there was a significant interaction between *Identical Cognate* and *English Secondary Family Size*, showing inhibition for identical cognates with increasing English secondary family size. The model did reveal a processing advantage for cognates in comparison to non-cognate controls. This facilitation effect was exclusively carried by the identical cognates: There was no significant difference between non-identical cognates and controls (hence the inclusion of *Identical Cognate* in the final model rather than *Cognate*). Finally, *PreviousRT* and *OLD* emerged as significant predictors of response latencies. The inhibitory effect of *PreviousRT* shows that items are responded to slower when the response latency of the preceding word item is long, while the inhibitory effect of *OLD* reveals a processing disadvantage for words with many close orthographic neighbors[i]. Finally, the positive correlation parameter for the by-subject random intercepts and random slopes for *Identical Cognate* indicate that slower participants responded less quickly to identical cognates.

Discussion

The results of Experiment 2 replicate the monolingual pattern observed in Experiment 1 with respect to both English primary and secondary family size. *English Primary Family Size* had a facilitatory effect on response latencies. This result extends the observed English primary family size effects in Dutch-English bilinguals of Dijkstra et al. (2005) on the processing of Dutch-English interlingual homographs in English lexical decision to the situation of cognates. Importantly, this effect shows that the bilinguals in our study were sensitive to morphological and semantic relationships for these words in their L2 and that they are sensitive to the morphological productivity of these L2 words during reading. There was no indication that English primary family size effects varied with the degree of form overlap with Dutch words, since no significant interaction between *English Primary Family Size* and either *Cognate* (cognates versus non-cognates) or *Identical Cognate* (identical versus other items) was observed.

Further, as expected, the bilinguals were sensitive to the cognate status of the stimuli. A cognate facilitation effect was observed that was entirely driven by the identical cognates and was absent for non-identical cognates. This dissociation between identical and non-identical cognates is in line with predictions made by localist connectionist models like BIA+ that predict a gradual decrease in response latencies with an increase in similarity for non-identical cognates and a steep decline in response latencies going from non-identical to identical cognates. This prediction was confirmed by bilingual lexical decision data of Dijkstra, et al. (2010; see also Van Assche, Duyk, Hartsuiker, & Diependaele, 2009). However, it should be noted that more than two-thirds of our non-identical cognates differed on two or three letter positions (e.g. *toma**to*** – *toma**at***). This suggests that the amount of overlap in these non-identical cognates may have been too small to trigger a cognate facilitation effect for these items.

Importantly, similar to what was observed in the monolingual data, there was a significant interaction between *Identical Cognate* and *English Secondary Family Size*, revealing longer response latencies for identical cognates with a large secondary family size. This shows that, even though bilinguals process words in their non-dominant language, they are sensitive to a larger chain of morphological relations, going beyond the primary family size. The finding that the facilitation for identical cognates relative to non-identical cognates and controls was attenuated for identical cognates with a large secondary family size can be explained by assuming a semantic origin of family size effects. The activated secondary family members of identical cognates are semantically unrelated to their target, and hence, constitute activated semantic noise. When the secondary family of an identical cognate is large, slower responses are produced relative to identical cognates that activate less semantically incongruent information.

Again, similar to what was observed in the monolingual data, the question arises of why the secondary family size effect is only observed for identical cognates and not in non-identical cognates and controls. Anticipating the results of our computational modeling to be discussed below, we will argue that the observed interaction between secondary family size and cognate status is a consequence of the distributional properties of English.

Interestingly, no effects of *Dutch Primary Family Size* and *Dutch Secondary Family Size* were observed. This could be due to the fact that in this experiment, the English family took away part of the effect of Dutch family size ('the winner takes it all'). We argue that cross-language family size effects are likely to be found in a paradigm in which the family size of the target language is kept constant, and in which the family size of the non-target language is contrasted. A recent study by Mulder, Schreuder, and Dijkstra (2012) on cross-language family size effects using behavioural and ERP measures indeed showed these cross-language effects in lexical decision on cognates when the family size of the target language was kept constant[ii].

The data of Experiment 1 and Experiment 2 was further analysed by means of a GAM[iii] (Generalized additive mixed model (Wood, 2006); see the Appendix for this analysis), using the same random effects structure as for Experiment 2, but with the additional predictor *AoA*. This joint analysis supported the presence of an effect of *OLD* in the second but not the first experiment (t = -2.9). It also supported a reduction in the magnitude of the effect of *Identical Cognate* for the monolinguals (t = 3.8). However, with increased power, the main effect of *Identical Cognate* reached significance (t = -5.2), indicating that, surprisingly, identical cognates may have a processing advantage even for monolinguals. The interaction of *Identical Cognate* by *Secondary Family Size* (t = 3.1) was not modulated further by an interaction with *Language* (monolingual/bilingual), indicating that across both experiments, the magnitude of the effect of S*econdary Family Size* was highly similar, and restricted to identical cognates. The joint analysis further revealed that monolinguals responded more quickly than bilinguals (t = -6.31), and that the effect of *English Frequency* was stronger for the bilinguals (t = 3.7). A similar reduction in the magnitude of the frequency effect as a function of response speed was observed by Baayen and Milin (2010) within a monolingual context across subjects. Finally, the joint analysis reveals a non-linear effect of *AoA (see Figure 1 of the Appendix)*, with slower RTs for words with a high *AoA* and faster RTs for words with a low *AoA*, and no effect in the middle range of the graph. Importantly, the effect of *AoA* only occurs for monolinguals, and not for bilinguals, disconfirming the suggested explanation of the effect of cognate status in terms of *AoA*.

In the Introduction, we asked whether the observed English family size effects are due to the resonance of activation between family members and targets in the lexicon, or whether these effects can be explained by more general learning principles applied to speakers' experience with the words of their language. In the following section, we will first discuss how interactive activation models account for the observed effects. Then, we present an alternative explanation in terms of computational simulations of the data of Experiments 1 and 2 with a model that works with just a single forward pass of activation, naïve discrimination learning.

Simulation study

Within the framework of spreading activation, the MFRM model (Morphological Family Resonance Model; De Jong, Schreuder, & Baayen, 2003) was a first attempt to specifically model family size effects. This monolingual interactive activation model explains family size effects by means of resonance between lemmas (see also Schreuder & Baayen, 1995) and the semantic and syntactic representations to which these lemmas are linked. When a semantic representation of a target word is linked to many associated lemmas (primary family members), a large amount of activation spreads back and forth between this semantic representation and the associated lemmas, gradually increasing the shared semantic activation and the activation level of the target lemma. Such resonance within the morphological family will thus amplify the rate at which the activation of the target lemma increases, speeding up recognition.

While this assumption of resonance of the model can account for the observed facilitation effect of primary family size, it cannot account for the inhibitory effect of secondary family size. Baayen (2010b) argued that this inhibitory effect arises because secondary family members generally activate semantic representations that do not overlap with that of the target word, under the assumption that lexical decision involves discrimination between semantically relevant and irrelevant meanings. Thus, activation of secondary family members such as *horse power* does not lead to faster responses to the target *work*, because their activated meaning will not strengthen the activation level of the target but rather compete with it. In interactive activation models, such as MFRM, resonance between morphological family members will always lead to facilitatory effects of family size. The MFRM fails to predict the inhibition from the secondary family size, and also fails to provide an indication of why this effect would be restricted to identical cognates.

For interactive activation models, there are two assumptions that must be made in order to make the right predictions. The first assumption is that identical cognates are characterized by two morphemic representations (rather than one), which are connected by inhibitory links. Recent evidence on French-English orthographically identical cognates from Peeters, Dijkstra, and Grainger (2013) suggests that this is a viable possibility for identical cognates. By adding inhibitory links between identical cognates, and by removing the links between non-identical cognates and control translation equivalents, the observed pattern of results (inhibition for identical cognates, no facilitation from secondary family size elsewhere) can be obtained. The second assumption lies in considering a task-decision system that can base its decisions on subsets of the activated representations, for instance, only on the basis of those semantic representations that are directly compatible with the target word. This suggestion would be in line with electrophysiological evidence from Mulder et al. (2012), who argue that ERP effects for family size are different from ERP effects for orthographic neighborhood size and associative neighborhood size because of their semantic overlap with the target word.

Instead of explaining the effects of primary and secondary family size in terms of interactive activation and task-decision level effects, in this paper, we can ask whether these effects can also be understood as a consequence of discrimination learning. Baayen et al. (2011) proposed a model, the naïve discriminative reader (NDR), that is a simple two-layer network with as (localist) input units letter unigrams and bigrams, and as (localist) output units, lexical meanings. In this model, there is a single forward pass of activation, from the input units to the output units. The model is a decompositional model in the sense that complex words and phrases are decomposed at the semantic level into the meanings of their constituents (e.g., *tea trolley* into *tea* and *trolley*).

The activation of a simple, mono-morphemic, word's meaning is obtained by summation over the weights from its letter unigrams and bigrams to its meaning. The activation of complex words and word n-grams is obtained by summation over the activations of the component meanings. Reaction times in the visual lexical decision task are modeled as inversely proportional to this (summed) activation. The model does not posit any separate representations for morphemes, complex words, or phrases. Nevertheless, it correctly captures whole word frequency effects, stem frequency effects, and phrase frequency effects (see Baayen, Hendrix & Ramscar, 2013). The model is theoretically anchored in the theory of discrimination learning

(Wagner & Rescorla, 1972; Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010), as formalized by the Rescorla-Wagner equations (see Appendix).

These equations, which formalize a substantial body of research on animal and human learning, characterize the strength of the association of a cue to an outcome as a complex dynamic system, the behaviour of which changes over time as a function of past experience. The association strengths between cues and outcomes increase or decrease depending on how well the cues predict a given outcome. The magnitude of the changes in association strength for a given cue and outcome are smaller when there are more cues present at a learning trial. The NDR model actually estimates the association strengths (weights) of cues (letters and letter bigrams) to outcomes (meanings) by means of the equilibrium equations for the Rescorla-Wagner equations derived by Danks (2003), obviating the need to simulate the learning process step by step. This opens the way for efficient estimation of the weights directly from large corpora.

It is worth noting that the weights are completely and exclusively determined by the distributional properties of the input. In other words, estimation of the weights is deterministic given the model input, typically words (or word n-grams) and their frequency of occurrence in a corpus or lexical database. For monomorphemic words, such as the words examined in the present study, the estimated activation of a given word's meaning proceeds without the intervention of free parameters. The activation of a word's meaning is completely and exclusively determined by the weights from that word's letter unigrams and bigrams to its meaning, which in turn are determined completely and exclusively by the corpus from which the weights are estimated.

The NDR model differs in several aspects from connectionist models such as the triangle model of Harm and Seidenberg (2004). First, the triangle model is more comprehensive, as it models the relation between orthography and pronunciation. The NDR in its current implementation therefore offers an implementation of only a part of a much richer cognitive system. Second, the NDR model is a localist model that does not make use of hidden layers, and it does not seek to understand higher-order generalizations in terms of patterns of activation over hidden units (see, e.g. Elman, 1990; McClelland & Rumelhart, 1986). Third, the NDR model learns from `raw' language data; no transformations of frequency such as used by the triangle model (equation 6) of Harm and Seidenberg (2004) are required. The NDR model has in common with the triangle model that it seeks to understand lexical processing without positing hierarchies of discrete form units for morphemes and words mediating the mapping from letter sequences to meaning.

The primary family size effect arises in the NDR model because a word's morphological family members provide a consistent learning environment that helps strengthen the weights from the word's letter unigrams and bigrams to its meaning. For instance, *teapot* and *teasing* both contain the orthographic string *tea*. In the case of *teapot*, the model strengthens the weights from the unigrams and bigrams of *tea* to the meaning ‘tea’, whereas in the case of *teasing*, the weights to ‘tea’ are decreased. The greater the number of family members, the stronger the weights from the letter unigrams and bigrams to ‘tea’ become.

Understanding the effect of secondary family size is less straightforward. For compounds, Baayen (2010b) observed complex non-linear interactions of secondary family size with head family size and membership of the strongly connected component of the English compound graph. Only a partial explanation of the secondary family size was presented, based on the observation that the orthographic similarity of modifier and head co-varied with the predictors in the interaction.

For monomorphemic words, the effect of family size has not been studied within the framework of naïve discrimination learning. Furthermore, the explanations suggested for compounds do not carry over to simple, monomorphemic, words. If the NDR correctly predicts a secondary family size effect for the words used in Experiments 1 and 2, then this would support the hypothesis that the effect arises due to the distributional properties of the words in the language in interaction with discrimination learning.

Simulation Experiment 1

For Experiment 1, a naïve discrimination network was set up for 27049 orthographically distinct lemmas with up to 10 letters from the CELEX lexical database, which jointly represent 18.1 million word tokens. The weights from the 721 letter unigrams and bigrams to the 16539 different constituent meanings were estimated using the equilibrium equations of Danks (2003), using the *ndl* package of Arppe, Milin, Hendrix, and Baayen (2011). The activations of the word meanings were obtained by summation over the weights from the letter unigrams and bigrams in the orthographic input to these meanings. Excellent results are already obtained when simulated reaction times are defined as minus the logarithm of the activations. The logarithmic transform, required to facilitate the statistical analysis, removed most of the skew from the distribution of activations, and the change of sign is motivated by the straightforward consideration that words that have been learned better (greater activation) can be responded to faster (shorter latency). Slightly improved results ensue when not only the target word's activation is taken into account, but also the summed activations of competitors, which is expected to speed responses (cf. the multiple read-out model of Grainger & Jacobs, 1996). To this end, we estimated from the data an activation threshold $\theta = 0.092$ such that the summed activation of all meanings (except the target meaning) above this threshold correlated maximally with the observed by-item mean response latencies. The resulting activation, α_θ, is a second predictor of the response latencies, along with the activation of the target meaning α_{target}. In order to estimate the relative weight of these two predictors, we made use of a linear model regressing observed reaction time on α_{target} and α_θ,

$$\text{log observed RT} \sim \beta_0 + \beta_1 \log(\alpha_{\text{target}}) + \beta_2\, \alpha_\theta \qquad (1)$$

resulting in the estimates -0.0124 for β_1 and -0.04577 for β_2 (both $p < .05$). Simulated reaction times were defined as the fitted values of this regression model.

In order to compare the simulated latencies with the observed latencies, we calculated mean RTs for Experiment 1, which were also log-transformed. The correlation between the observed and simulated reaction times was 0.32 ($t(92) = 3.20$; $p = 0.0019$).

In order to evaluate the extent to which effect sizes are comparable for the observed and simulated reaction times, we regressed the simulated latencies on the predictors that reached significance in the analysis of the observed latencies in Experiment 1: word frequency, primary family size count, secondary family size, cognate status (identical, non-identical, control) and cognate status by secondary family size. Figure 4 plots the coefficients (excluding the intercept) of the model fitted to the simulated latencies on the horizontal axis, and the coefficients of the model fitted to the observed latencies on the vertical axis. Table 4 presents the coefficients of the model fitted to the simulated latencies along with their corresponding *t*-value and *p*-value.

(Figure 4 and Table 4 about here)

The correlation of the two sets of coefficients was 0.95 ($t(5) = 6.84$; $p = 0.0010$). With just 94 items, only the coefficients of frequency and family size reached significance for the simulated latencies. However, the relative effect sizes are estimated accurately, which indicates that the effects of frequency, primary and secondary family size, as well as cognate status, can all be understood as arising in a dynamic system based on simple and well-understood principles of learning that is exposed to the distributional properties of English form to meaning mappings.

It is worth noting that virtually the same results are achieved by a model that has no free parameters whatsoever, i.e., by a model that takes only the activation of the target meaning into account. The full model, however, fits well with earlier work on multiple-readout of evidence for lexicality. The present model shows that the insights originally formulated within the interactive activation framework can be integrated within the framework of naïve discrimination learning.

To see why an effect of secondary family size arises in the model, we first call attention to the pervasive role of compounding in structuring the English lexicon. Compounding is the most productive word formation process in English, and most familial ties are carried by compounds. For instance, *tea* and *bus* are secondary family members through a morphological chain carried by two compounds, *tea-trolley* and *trolley bus*. The secondary family size effect hinges on the links in such chains, in the present example, *trolley*. When *trolley* co-occurs with *tea*, the weights from its unigrams and bigrams to the meaning 'tea' are strengthened. Whenever *trolley* occurs in *trolley bus*, the weights from *trolley* to 'tea' decrease and those for 'bus' are strengthened.

More specifically, the weights of *trolley* to 'tea' co-determine the weights of *tea* to 'tea' through the sums in the Rescorla-Wagner equations $\sum_{\mathrm{PRESENT}(C_{j,t})} V_j$ in equation (3) of the

Appendix. When *trolley* occurs in few other compounds, the letter unigrams and bigrams of *trolley* will contribute little to these sums for the outcome 'tea', other things being equal. As a consequence, the change in the weights on the connections from the letter unigrams and bigrams of *tea* to the meaning 'tea' will not be affected much. However, when *trolley* occurs in many other compounds, and develops negative weights to 'tea', then the connection weights of *tea* to 'tea' will be adversely affected. With reduced weights, activations decrease, and hence simulated RTs for 'tea' increase.

We cannot offer a detailed explanation, however, of why the effect of secondary family size is restricted to the identical cognates, both for monolingual speakers of English, for the simulation, and, as will become apparent below, for Dutch-English bilinguals. Apparently, the co-occurrence patterns of orthographic cues and meanings in English are such that in the course of learning, identical cognates acquire a processing advantage that decreases with increasing secondary family size.

Simulation Experiment 2

For the modeling of Experiment 2, we explored two different modeling strategies. The first strategy pursues the idea that the experience with Dutch and English is completely merged into a single unified network. The second strategy explores the possibility that Dutch and English have separate networks that are accessed in parallel. Both strategies make use of the same English instance base as was used for Experiment 1, complemented by a Dutch instance base that we also derived from CELEX. As for English, only lemmata with less than 11 letters were included, resulting in an instance base with 29802 unique lemmata representing 33.7 million word tokens, and comprising 9486 different constituent meanings.

When it is assumed that English is integrated into the network of Dutch (strategy 1), the weights are calculated from the combined Dutch and English instance bases. Within this joint instance base, we assigned the same meaning representations to the identical and non-identical cognates in both languages. We defined simulated latencies as minus the log of the activation (-log $\alpha_{\text{target})}$, as in the simulation study of Experiment 1, resulting in a correlation with the observed latencies of 0.29 ($t(90) = 2.929$; $p = 0.0043$).

For the bilingual latencies, further inspection indicated a multiple read-out approach to improve results, as was the case for Experiment 1**.** The summed activation of meanings other than the targeted meanings exceeding an activation threshold of 0.31 turned out to co-predict the observed response latencies in a linear model regressing observed RT on -log(a) ($\beta = 0.025$; $p = 0.0004$) and the activation α_θ exceeding the activation threshold ($\beta = -0.078$; $p = 0.0084$).

The activation α_θ was orthogonal to the lexical predictors, and captures subjects' response strategies. It was estimated from the data by regressing the observed RTs on α_θ for a range of thresholds and selecting that threshold value for which the largest (negative) correlation was observed.

We then regressed α_θ out of the observed RTs. The model regressing the denoised RTs on the lexical predictors provided a slightly better fit (the AIC improved from -232 to -237). The correlation of the denoised RTs and the activations of the meanings -log(a) was 0.35 ($t(90) = 3.51$; $p = 0.0007$). Note that as a consequence of this denoising, the model for the Dutch-English bilinguals has two free parameters, namely, the intercept and slope used to regress α_θ out of the observed RTs.

Next, we examined whether the relative effect sizes for the simulated latencies resemble the effect sizes for the observed latencies. We used the same model specification as for

Experiment 1, regressing the simulated latencies on word frequency, primary and secondary family size, cognate status (identical, non-identical, control), cognate status by secondary family size, and OLD. Coefficients for observed and simulated latencies were highly correlated (r = 0.835; p = 0.0099). Table 5 presents the coefficients of the model fitted to the simulated latencies along with their corresponding *t*-value and *p*-value.

(Figure 5 and Table 5 about here)

However, Figure 5 clarifies that the effect size of secondary family size for identical cognates status is much too small. This may in part be due to a non-optimal coding of translation equivalents in the morphological families of the two languages. Working with this model, however, leads us to think that the Dutch system in this bilingual model is acting as a source of noise masking the effect of the English system that was visible for the monolinguals.

We therefore also explored strategy 2, according to which Dutch and English are learned in two separate networks. When a word is read, its orthographic cues (letter unigrams and bigrams) are activated. These cues activate meanings in both networks. For a given input, say *frog* ('kikker'), with orthographic cues (*f, r, o, g, #f, fr, ro, og, g#*), the activation of the meaning 'frog' is calculated for English, by summation over the weights from the cues to the meaning in the English lexicon, resulting in the English activation α_E. The activation of the corresponding meaning in Dutch, α_D, was obtained in the same way. Note that strategy 2 remains compatible with the hypothesis of non-selective access, as both networks are accessed in parallel.

For each network, we calculated an activation threshold, such that the summed activation of non-targeted meanings with activations exceeding this threshold correlated maximally with the response latencies. The summed activation for Dutch, $\alpha_{\theta,D}$, turned out to be a significant predictor of the response latencies. This was not the case for the summed activation for English**,** however. Log-transformed simulated reaction times were defined as

$$\text{log simulated RT} = -0.0107 \log(\alpha_D) - 0.01903 \log(\alpha_E) + 0.0291\, \alpha_{\theta,D}. \qquad (2)$$

The three weights, the free parameters of this model, were obtained by means of the linear regression model

$$\text{log observed RT} \sim \beta_0 + \beta_1 \log(\alpha_D) + \beta_2 \log(\alpha_E) + \beta_3\, \alpha_{\theta,D}. \qquad (3)$$

The correlation between the by-item observed and simulated reaction times was 0.33 ($t(90) = 3.37$, $p = 0.0011$), indicating a good fit at the item level.

Interestingly, the coefficient of $\alpha_{\theta,D}$ was positive, indicating that Dutch-English participants doing lexical decision in English are slowed by the activation of inappropriately activated meanings in their mother tongue. The positive slope of $\alpha_{\theta,D}$ for bilinguals contrasts with the negative slope of the corresponding activation for monolinguals.

For evaluating goodness of fit at the level of effect sizes, we inspected the correlation between the coefficients of the regression models fitted to the observed and expected RTs, which indicated a satisfactory fit ($r = 0.93$, $t(6) = 6.28$, $p = 0.0008$, see Figure 6). Furthermore, those and only those coefficients that reached significance for the observed latencies also reached significance (all $p < 0.10$, i.e., significant in the expected direction) in the model for the simulated latencies. Table 6 presents the coefficients of the model fitted to the simulated latencies along with their corresponding *t*-value and *p*-value.

(Figure 6 and Table 6 about here)

Strategy 2 clearly leads to a superior model, although at the price of one additional free parameter, and a more complex network structure. The improved results indicate that the Dutch and English networks are likely to be subject to domain-specific learning. However, the simulations with the NDR are based on task-specific data of a particular target language. It can therefore not be excluded that task-specific mechanisms have affected learning. More simulations are needed to clarify this.

In summary, naïve discrimination learning is successful in accounting for primary and secondary family size effects in both monolingual and bilingual processing. Importantly, the NDR model reproduces the interaction between secondary family size and identical cognate status observed across both experiments. Furthermore, it also captures the processing advantage of identical cognates, even for monolinguals (an effect that the joint analysis of both experiments revealed to be robust across the two groups of participants). The good fits obtained indicate that the effects of cognate status and family size (both primary and secondary) can be understood as arising from a simple learning system (as defined by the Rescorla-Wagner equations) that is exposed to language use. It is also worth noting that a joint analysis of the simulated latencies for Experiment 1 and Experiment 2 reveals a significant interaction of word frequency by language, with a reduced frequency effect for monolinguals (p = 0.04), replicating the same interaction for the observed latencies.

Distributional analysis of data from the English Lexicon Project

Modeling with naive discrimination learning suggests that the distributional properties of the English lexicon underlie the effect of secondary family size for identical cognates. We therefore investigated whether cognate (identical or non-identical) versus non-cognate status is predictable from the lexical-distributional predictors available to us. Interestingly, the one predictor on which cognates and non-cognates were not matched, secondary family size, is predictive, such that a higher secondary family size raises the probability of a word falling into the non-cognate class $\hat{\beta}$= 0.30,t = 3.075, p = 0.0022$). This suggests that within the present sample from the English lexicon, cognates are found in less dense regions of the lexical graph. Within the set of cognates, secondary family size does not differentiate further between identical and non-identical cognates. This suggests that the naive discrimination model detects subtle aspects of the form-meaning mapping that are beyond a crude connectivity measure such as the secondary family size count. Nevertheless, it is clear that cognate status is not orthogonal to the distributional properties of the lexicon.

Further support for this conclusion is provided by a survey of a second, independent, sample of cognates. Since only 6 of the 25 identical cognates in Experiment 1 and 2 were monosyllabic, we selected as materials the monosyllabic, monomorphemic words from the English Lexicon Project (henceforth ELP; Balota et al., 1999, 2007), that were previously studied in detail in Baayen et al. (2006). Each of the words (collapsing into one word type those conversion alternants that can be either a noun or a verb) was inspected and evaluated as an identical cognate if the word exists with exactly the same spelling in both English and Dutch, and if the word has at least one meaning in common in both languages. E.g., *arm* in Dutch can mean both the human upper limb but also `poor', whereas in English, *arm* denotes the upper limb, but can also mean `to supply with weapons'. Since the meaning `limb' is share by both languages, *arm* was classified as an identical cognate. Whether words share a historical ancestor in the etymological sense was not used as a criterion. The 2197 words in the data set comprised 224 identical cognates, 554 non-identical cognates, and 1419 words that were unrelated.

An advantage of inspecting monosyllabic and monomorphemic words is that here, neighborhood density and orthographic consistency measures are most precise. For a three-syllable word such as *alcohol*, the a-priori chances of having an orthographic neighbor at a Hamming distance[iv] of one is much smaller than for a monosyllabic word such as *tent*. By focusing on words with highly restricted phonotactic and orthotactic structure, chances are optimized for detecting possible relations between cognate status and phonological and orthographic similarity.

We opted for using the lexical decision data from the American megastudy (Balota et al., 2007) instead of its British counterpart (Keuleers, Lacey, Rastle, Brysbaert, 2012) because in the US, speakers are more likely to be truly monolingual than in the UK. In the UK, children are more likely to receive at least some education in a foreign language.

Table 7 and Figure 7 present the results of a logistic regression analysis predicting the log odds of a word being an identical cognate. (Results for models predicting cognate versus non-cognate status are similar, and are not reported.) Longer words are less likely to be identical cognates, and the same holds for verbs as opposed to nouns. Furthermore, words beginning with a voiceless segment have a greater probability of being an identical cognate.

(Table 7 and Figure 7 about here)

Two orthogonal measures of neighborhood density and orthographic consistency taken from Baayen et al. (2006) also reached significance. These measures, *PC1* and *PC2,* are latent variables constructed with principal components analysis from 10 highly collinear measures of orthographic and phonological neighborhood density and consistency.

PC1 contrasts contrasted forward enemies (number of words with different pronunciation for the same sequence of letters), with small positive loadings, with phonological and orthographic neighbors (number of words that differ by a single phoneme or letter) and friends (words with the same letter sequence and the same pronunciation), with large positive loadings. This variable is correlated with the N-count neighborhood density measure, $r = 0.357$ ($t(2195) = 17.91$, $p < 0.0001$), but outperforms this measure in all analyses reported below. The effect of *PC1* was linear, and indicated that words with many neighbors and more frequent neighbors as well as more friends (both orthographic and phonological) are less likely to be identical cognates.

PC2 contrasts friends and spelling neighbors, which have positive loadings, with backward enemies (words with the same pronunciation but a different spelling) and phonological neighbors, which have negative loadings. *PC2* entered into an interaction with log-transformed written frequency in the British National Corpus. The partial effect of this interaction is shown in Figure 7. Across most of the range of *PC2*, the effect of frequency is inverse U-shaped. The effect of frequency reaches its peak amplitude for the highest values of *PC2*. In other words, the probability that a word is an identical cognate is greatest for words that enjoy medium to high frequency of use and that have many neighbors and many friends, but few backward enemies and few phonological neighbors (across both type and token-based counts).

Other predictors discussed in Baayen et al. (2006) and available in the *english* dataset in the *languageR* package (Baayen, 2010a) did not reach significance (including the classical N-count measure for neighborhood density).

In summary, monosyllabic identical cognates are less likely to have many orthographic and phonological neighbours (*PC1)*, and when they do have many orthographic (but not phonological) neighbors, they are protected by a high frequency of use. Furthermore, high-frequency words are unlikely to be identical cognates when they have many backward enemies and many phonological neighbors. Identical cognates thus appear to inhabit an `ecological niche' in lexical distributional space where they are orthographically unique expressions of the phonology, with few phonological neighbors, and protected by frequency against loss of discrimination against orthographic neighbors.

Finally, we consider the possibility that lexical decision latencies in the ELP are co-determined by cognate status. A GAM fitted to the log-transformed lexical decision latencies of the young subjects in the ELP supports this possibility. Table 8 summarizes the full model, which improves on the model reported previously in Baayen et al. (2006). (The model also supports an effect of secondary family size, in interaction with primary family size.) Visual inspection of the smooths for frequency indicates that for monosyllabic words, the frequency effect is stronger for cognates than for non-cognates, especially in the lower-frequency range. From the logistic model predicting cognate status, we know that low-frequency cognates tend to be words with many backward enemies and many phonological neighbors, i.e., words with increased phonological uncertainty. It is conceivable that this increased phonological uncertainty concerning a word's pronunciation is responsible for the elongated response latencies to cognates compared to non-cognates.

(Table 8 about here)

For Experiment~1 and~2, we observed that identical cognates were responded to faster compared to non-identical cognates and non-cognates, which raises the question why lower-frequency cognates (including non-identical cognates, model not shown) elicit longer response latencies than non-cognates in the ELP data set. We think the reason is that the critical identical cognates in Experiment 1 and 2 have few backward enemies (cf. *alcohol*, *camera*, *sultan*, *horizon*, *toilet*, *hotel*, *minister*, *opera*, *ego*) and few phonological neighbors. As a consequence, they are not subject to the costs of phonological uncertainty, and can be responded to more quickly, exactly as predicted from the orthography by our NDR model.

These analyses allow us to conclude that cognates do not enjoy a special status in that they would be `flagged' for special treatment in lexical processing. Instead, cognates occupy ecological niches where they maintain an orthography that tends to be paired with a phonology without many neighbors. Short low-frequency cognates, however, tend to come with more phonological uncertainty, and a concomitant processing cost in lexical decision.

General Discussion

The aim of this paper was to investigate the co-activation of lexical representations in the bilingual mental lexicon. Lexical representations can be related in many ways. In terms of their orthography or/and phonology, lexical representations might share a part or even their complete form (e.g., the English word *book* and Dutch word *boek*). At the semantic level, lexical representations might overlap in meaning (e.g., the English word *bicycle* and Dutch word *fiets*). When there is overlap in both form and meaning, lexical representations might also be related in terms of their morphology (e.g., the English words *book* and *bookcase* and, the Dutch words *boek* and *boekenkast*). In this paper, we have explored the degree to which these different forms of relationships play a role in bilingual word processing. We addressed this issue by looking at primary and secondary L2 family size effects (due to morphological and semantic overlap) on the processing of cognates by Dutch-English bilinguals.

We first tested English monolinguals on the selected stimulus materials with more precise primary family size counts in a lexical decision task (Experiment 1). These family size counts were generated by bilingual Dutch-English participants. Given the high correlation between family size counts for the cognate items generated by the monolinguals and those generated by the bilinguals, the new family size measure can be used with confidence to assess family size effects in both data sets. Moreover, applying the same measure to both dataset increases the comparability of effects.

The new primary family size measure turned out to be a better predictor than the original CELEX family size counts. An overall facilitatory effect of primary family size was observed; a secondary family size effect was observed for identical cognates only. A higher English secondary family size led to inhibition for identical cognates.

In the data for Dutch bilinguals, facilitatory English primary family size effects were observed for both cognates and English control items. These results demonstrate that Dutch bilinguals are sensitive to the primary morphological productivity of L2 words, extending the results of Dijkstra et al.'s (2005) study on interlingual homographs (e.g., words that share their form but not their meaning in two languages) to the situation of cognates. Dijkstra et al. observed facilitatory effects of the primary family size of the target language in both English and Dutch lexical decision. Our study replicated this effect for cognates.

Further, an important finding of our study is that Dutch-English bilinguals were even sensitive to the secondary family size of words of their L2. Similar to what was observed for the English monolinguals in Experiment 1, a higher English secondary family size slowed down the processing of identical cognates. An inhibitory effect of secondary family size fits well within a semantic explanation of the family size effect as proposed by Mulder et al. (2012), and outlined in the Introduction.

The finding that secondary family size only affected the processing of identical cognates in both the bilingual and monolingual data was not expected. Though the direction of the secondary family size effect (i.e., inhibition) is in accordance with the effect of secondary family size observed by Baayen (2010b) for English two-constituent compounds with small head primary family sizes, it is not clear why there was no effect for English (mono-morphemic) control words or English-Dutch non-identical cognates in either the monolingual or the bilingual data.

We initially argued that, in bilinguals, secondary family size effects are more likely to affect the processing of cognates, and, specifically, identical cognates, rather than English control words. The underlying reason for this assumption is that identical cognates may have linked representations in the bilingual mental lexicon due to their formal overlap with words in their dominant language, and, consequently, the subsequent co-activation of items in both languages would facilitate the spreading of activation to the primary and secondary family members of the L2. In support of this, Experiment 2 indeed showed that English (L2) secondary family size only affects identical cognates but not non-identical cognates and English controls.

However, the finding that secondary family size only affects identical cognates in the monolingual data as well does not support this argument. Moreover, a further surprising result revealed by the more powerful omnibus analysis of both experiments was a significant processing advantage for identical cognates not only for bilinguals but also for monolinguals. This suggests that the observed effect of English secondary family size for identical cognates in the bilingual data are unlikely to be a consequence of a facilitated spreading of activation due to the co-activation of items in the non-target language and therefore an explanation for this effect should be sought elsewhere.

Although the facilitatory effect of English primary family size can be accounted for by interactive activation models such as MFRM, spreading activation alone cannot explain the observed inhibitory effects of English secondary family size nor why it only affects identical cognates. Without additional assumptions, spreading of activation between morphological family members in interactive activation models will always lead to facilitatory effects. The effects observed in Experiments 1 and 2 show that resonance of activation between indirectly related lexical items in the lexicon cannot be the mechanism underlying the slower responses to words with a larger number of secondary family members. It is worth noting that the secondary family size effect seems to challenge a simple multiple read-out mechanism (Grainger & Jacobs, 1996), according to which a lexical decision can be facilitated when many competitors are highly activated. Under such an account, one would expect secondary family members to facilitate lexicality decisions, instead of inhibition.

As we argued above, this problem with MFRM and interactive activation models more generally can be resolved in at least two ways. First, by assuming that identical cognates have two, mutually inhibiting, morphemic representations (Peeters et al., 2013). Second, in more complex interactive activation models like BIA+ (Dijkstra & Van Heuven, 2002), a task-decision system could be involved that can make task- and trial-dependent decisions by basing itself flexibly on multiple information sources (also see Mulder et al., 2012).

Alternatively, the results of Experiments 1 and 2 were simulated successfully by the naïve discriminative reader model, which replicated the critical interaction of identical cognate status by secondary family size. These simulations provide an alternative account for how family size effects arise, and the differences between interactive activation models and the NDR model indicate that they are not completely functionally isomorphic. Instead of being seen as a consequence of activation spreading in a network of lexical nodes, they are understood as a consequence of the process of learning to map orthographic input onto meanings. The weights on the connections evolve during learning to optimally discriminate between different meanings, given the distributional properties of the language and its writing system to which the learner is exposed[v]. In this dynamic systems approach, it is found that primary family members tend to facilitate learning, whereas secondary family members appear to render learning more difficult. As a consequence, response latencies in the visual lexical decision task are shorter for words with large primary families, but longer for words with large secondary families. For the present data, our simulation studies strongly suggest that the inhibitory effect of secondary family size specifically for identical cognates is a consequence of how the distributional properties of English happen to fall out for identical cognates. This conclusion is further supported by the presented model predicting cognate status using the RT data from the English Lexicon Project.

When working with interactive activation models, the question arises whether family size effects arise as a consequence of activation spreading among word forms, or among word meanings. For the primary family size effect, there is a growing body of evidence, as discussed in the introduction, that word meanings are crucially involved. For the secondary family size effects, a semantic locus also seems more likely. As observed above, it is only a semantic account that straightforwardly predicts an inhibitory effect. In the introduction of this paper, we argued that the semantic (in)congruence between a target word and its family members determines whether facilitation (for semantically related meanings) or inhibition (for semantically unrelated words) is observed. In current interactive activation models, such as BIA or BIA+, the mapping between representations is based on purely formal (i.e., orthographic) information links. In contrast, the NDR model works with a direct mapping from orthographic cues to semantic outcomes. It is this direct mapping, crucially framed within the well-motivated learning regime of the Rescorla-Wagner equations, which enables it to account for effects of semantic (in)congruence, and as a consequence, for the observed primary and secondary family size effects .

Within the framework of naïve discrimination learning, the question of whether word forms or word meanings are at issue does not arise, as the model rejects morphemes and word forms as superfluous theoretical constructs. In this respect, the NDR model resembles the triangle model of Harm and Seidenberg (2004). Family size effects, both primary and secondary, are now an emergent property of a dynamic system learning the mapping of letters and letter bigrams to meanings.

Interestingly, the model that best fits the bilingual data is a model based on two separate networks that are accessed in parallel. It is important to note that, even though the model argues against the idea of a fully integrated bilingual lexicon, it is compatible with the hypothesis of language non-selective access (cf. discussion in Van Heuven et al., 1998), and indicates that the Dutch and English networks are subject to domain-specific learning. This architecture is consistent with the finding that associations between words within and between languages are not necessarily identical in L1 and L2 processing (Van Hell & De Groot, 1998). This supports the proposal that words such as cognates do not necessary have a fully shared representation in the lexicon but that part of their, at least semantic, representation is separate (cf. Peeters et al., 2013).

The simulation studies also integrated the notion of multiple read-out (Grainger & Jacobs, 1996) by including as a predictor the thresholded summed activation of competitors. For English monolinguals, this activation was facilitatory: Participants used this activation as evidence for a positive lexicality decision. For Dutch-English bilinguals, however, this activation, restricted to the Dutch network, was inhibitory, indicating that these participants found it difficult to suppress misleading information provided by their mother tongue. These results show that the naïve discriminative reader model can be extended with task-specific components, and illustrate the more general point that the learning network in this model is only a small part of a much richer cognitive system.

The simulation study in terms of naïve discrimination learning is insightful in several ways. First, it clearly shows that simulations by interactive activation models like MRFM and BIA+ may result in qualitatively problematic outcomes as long as parts of the network (e.g., the mapping orthography on semantics or the decision component) are not fully implemented. Simulations with more complete and complex frameworks like Multilink (Dijkstra & Rekké, 2010) are therefore in order. Second, the innovative study on discrimination learning presented here has focussed on structural issues (i.e., the mapping of orthography on semantics during learning) and has not considered how to simulate different patterns of results that follow from processing differences due to task demands. Models like BIA+ and the IC model (Green, 1998) explicitly include a task-decision system to account for systematic, task-dependent variability in empirical results across tasks. It remains to be seen how a naïve discrimination learning framework can be extended to include a rich task system.

In the present study, following the multiple read-out approach, we have made a first step by showing that a task component specific to lexical decision can be integrated in the NDR model, and that this integration results in a better fit to the observed response latencies. We think it is impressive to see how far this new localist framework can come with very simple assumptions, a minimum of free parameters, and full-scale corpus data.

Finally, the NDR model provides an intriguing new perspective on what a lexical network might look like. The intuitive and familiar representation of a lexical network formally is that of a graph with words as vertices and lexical (familial) relations as edges. In such a framework, the primary family size measure captures what in graph theory is called the edge degree of a vertex. The network of the NDR, by contrast, is much simpler in structure, with edges from orthography to meaning, but with no edges between semantic vertices. What the NDR shows is that nevertheless the Rescorla-Wagner learning principles allow a simple two-layer network to absorb in its weights many of the semantic properties that in the familiar interactive scenario take place between word vertices. The challenge for future research is to separate out effects that truly belong to the Rescorla-Wagner network learning to map form onto meaning, from effects that are a genuine part of the network of relations between the meanings themselves.

To summarize, our study is the first to investigate and model both primary and secondary family size effects in monolingual and bilingual word processing. After developing a more sensitive measure for English primary family size effects, we observed effects of both the primary and secondary family size for cognates in English visual lexical decision, for both monolinguals and bilinguals. The simulations were a first step to model primary and secondary family size effects in both monolingual and bilingual word processing within the framework of naïve discrimination learning. Whereas interactive activation models are challenged by the inhibitory effect of secondary family size for identical cognates, naïve discrimination learning provides an adequate account for the observed primary and secondary family size effects and the latter's interaction with cognate status. Our study shows that, despite a lower proficiency in English compared to monolinguals, Dutch bilinguals show the same surprising interaction of secondary family size and cognate status. Apparently, bilinguals are able to build lexical networks for their second language that are remarkably isomorphic with the networks of monolinguals.

Author note

Unfortunately, one of the coauthors, our friend and collegue, prof. dr. Rob Schreuder, passed away.

Acknowledgements

The authors wish to thank Walter van Heuven for providing the facilities to collect the monolingual data at the University of Nottingham.

References

Andrews, S. (1997). The effects of orthographic similarity on lexical retrieval: Resolving neighbourhood conflicts. *Psychonomic Bulletin & Review*, *4*, 439-461.

Arppe, A., Milin, P., Hendrix, P., & Baayen, R. H. (2011). ndl: Naïve discriminative learning [Computer software manual]. Available from http://CRAN.R-project.org/package=ndl (R package version 0.1.1)

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

Baayen, R.H. (2010a). *languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics*". R package version 1.0.

Baayen, R.H. (2010b). The directed compound graph of English. An exploration of lexical connectivity and its processing consequences. In S. Olson (ed.), *New impulses in word-formation (Linguistische Berichte Sonderheft 1)* (pp. 383-402), Buske, Hamburg.

Baayen, R. H. (2013). Learning from the Bible: Computational modeling of the costs of letter transpositions and letter exchanges in reading Classical Hebrew and Modern English. *Lingue & Linguaggio*, in press.

Baayen, R.H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Baayen, R.H., Feldman, L.F., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language, 53*, 496-512.

Baayen, R.H, Lieber, R., & Schreuder, R. (1997). The morphological complexity of simplex nouns. *Linguistics*, 35, 861-877.

Baayen, R.H., Hendrix, P. & Ramscar, M. (2012). Sidestepping the combinatorial explosion: Towards a processing model based on discriminative learning. To appear in *Language and Speech*.

Baayen, R. H., Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*, 12-28.

Baayen, R. H., Milin, P., Filipovic Durdjevic, D., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naïve discriminative learning. *Psychological Review*, *118*, 438-481.

Baayen, R.H., Piepenbrock, R., & Gulikers, (1995). *The CELEX lexical database [CD-ROM]*. University of Pennsylvania Linguistic Data Consortium.

Balota, D.A., Cortese, M., & Pilotti, M. (1999). Visual lexical decision latencies for 2906 words *[On-line], Available: http://www.artsci.wustl.edu/~dbalota/lexical_decision.html.*

Balota, D.A., & Lorch, R. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 336-345.

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchinson, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman T. (2007). The English Lexicon Project. *Behavior Research Methods, 3,* 445-459.

Bertram, R., Baayen, R.H., & Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language, 42*, 390-405.

Boudelaa, S., & Marslen-Wilson, W.D. (2011). Productivity and priming: Morphemic decomposition in Arabic, *Language and Cognitive Processes, 26*, 624-652.

Brysbaert, M., & New, B. (2009).Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977-990.

Chwilla, D.J., Kolk, H.J., & Mulder, G. (2000). Mediated priming in the lexical decision task: Evidence from event-related potentials and reaction time. *Journal of Memory and Language*, *42,* 314-341.

Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology*, *47*, 109-121.

De Groot, A.M.B.(1983). The range of automatic spreading activation in word priming.*Journal of Verbal Learning and Verbal Behaviour, 22,*417-436.

De Jong, N.H. (2002). *Morphological families in the mental lexicon.* MPI Series in Psycholinguistics, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

De Jong, N.H., Feldman, L.B., Schreuder, R., Pastizzo, M. & Baayen, R.H. (2002). The processing and representation of Dutch and English compounds: Peripheral morphological, and central orthographic effects. *Brain and Language*, *81*, 555-567.

De Jong, N. H., Schreuder, R., & Baayen, R.H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, *15*, 329-365.

De Jong, N. H., Schreuder, R., & Baayen, R.H. (2003). Morphological resonance in the mental lexicon. In R.H. Baayen & R. Schreuder (Eds.) *Morphological structure in language processing* (pp. 65-88). Berlin, Germany, Mouton de Gruyter.

Dijkstra, T. (2005). Bilingual visual word processing and lexical access. In J.F. Kroll & A. De Groot (Eds.) *Handbook of bilingualism: Psycholinguistic approaches* (pp.178-201). Oxford University Press.

Dijkstra, T., Miwa, K., Brummelhuis, B., Sapelli, M., & Baayen, R.H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, *62*, 284-301.

Dijkstra, T., Moscoso del Prado Martín, F., Schulpen, B., Schreuder, R., & Baayen, R.H. (2005). A roommate in cream: Morphological family size effects on interlingual homograph recognition. *Language and Cognitive Processes, 20*, 7-41.

Dijkstra, T., & Rekké, S. (2010). Towards a localist-connectionist model of word translation. *The Mental Lexicon, 5*, 401-420.

Dijkstra, T., & Van Heuven, W.J.B. (1998). The BIA model and bilingual word recognition. In. J. Grainger & A.M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189-225). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.

Feldman, L. B., & Siok, W. W. T. (1997). The role of component function in visual recognition of Chinese characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 776–781.

Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from Chronolex. *Frontiers in Psychology*, 2, 1-10.

Frost, R. (2012). Towards a universal model of reading. *Behavioural and Brain Sciences*, 35, 1-67.

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

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103* , 518-565.

Harm, M.W., & Seidenberg, M.S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662-720.

Juhasz, B.J., & Berkowitz, R.N. (2011). Effects of morphological families on English compound word recognition: A multitask investigation. *Language and Cognitive Processes*, *26*, 653-682.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*, 287-304.

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes, 23*, 1089-1132.

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 876-895.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English, *Behavior Research Methods*, *44*, 325-343.

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Sovjet Physics Doklady, 10*, 707-710.

Lüdeling, A., & De Jong, N.H. (2002). German particle verbs and word-formation. In N. Dehé, R. Jackendoff, A. McIntyre, & S. Urban (Eds.), *Verb-particle explorations* (pp. 315-333). Berlin, Germany: Mouton de Gruyter.

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*, 375-405.

McClelland, J.L., & Rumelhart, D.E. (1986). *Explorations in the microstructure of cognition, Vol. 2: Applications*, Cambridge, MA: Bradford.

McNamara, T.P., & Altarriba, J. (1988). Depth of spreading activation revisited: Semantic mediated priming occurs in lexical decision. *Journal of Memory and Language*, *22*, 545-559.

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Baayen, R. H. (2003). Morphological family size in a morphologically rich language: the case of Finnish compared to Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1271–1278.

Moscoso del Prado Martín, F., Deutsch, A., Frost, R., Schreuder, R., De Jong, N. H., & Baayen, R. H. (2005). Changing places: a cross-language perspective on frequency and family size in Hebrew and Dutch. *Journal of Memory and Language*, *53*, 496-512.

Mulder, K., Schreuder, R., & Dijkstra, T. (2012). Morphological family size in L1 and L2 processing: An electrophysiological study. *Language and Cognitive Processes*, *28*, 1-32.

Müller, O., Duñabeitia, J.A., & Carreiras, M. (2010). Orthographic and associative neighbourhood density effects: What is shared, what is different? *Psychophysiology*, 47, 455-466.

Peeters, D., Dijkstra, T., & Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language*, *68*, 315-332.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, *34*, 909-957.

Schepens, J., Dijkstra, T., van Heuven, W.J.B., & Grootjen, F. (2013). Cross-language distributions of high frequency and phonetically similar cognates. *PLOS ONE*.

Schreuder R., & Baayen, R.H. (1995). Modeling morphological processing. In Feldman, L.B. (Ed.), *Morphological Aspects of Language Processing*, pp. 131-154. Lawrence Erlbaum, Hillsdale, New Jersey.

Schreuder, R. & Baayen, R.H. (1997). How complex simplex words can be. *Journal of Memory and Language, 37*, 118-139.

Schulpen, B., Dijkstra, T., & Schriefers, H.J. (2003). *L2 proficiency and task effects on interlingual homograph recognition by Dutch-English bilinguals*. Unpublished manuscript, University of Nijmegen.

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change cognate effects in a sentence context? *Psychological Science, 20*, 923-927.

Van Hell, J.G., & De Groot, A.M.B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, *1*, 193-211.

Van Heuven, W.J.B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighbourhood effects in bilingual word recognition. *Journal of Memory and Language*, *39*, 458-483.

Wagner, A., & Rescorla, R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning ii* (pp. 64-99). New York: Appleton-Century-Crofts.

Wilson, M.D. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers*, 20, 6-11.

Wood, S. (2006).*Generalized Additive Models: An introduction with R*. CRC Press.

Yarkoni, T., Balota, D. & Yap, M. (2008). Moving beyond Coltheart's *N*: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*, 971-979.

Appendix

Formal definition of secondary family size

A formal definition of the secondary family size of a word ω (e.g., *horse* in Figure 1) proceeds as follows. Let F denote the set of bimorphemic words sharing ω as a constituent (the word with the constituents in light blue in Figure 1). This set includes all words with ω in first or in second constituent position. Let G denote the set of all words sharing at least one constituent with a word in F (all words with constituents that are colored in Figure 1; note that $F \subset G$). The secondary family size is defined as the cardinality of the set of words S which contains all words in G that are not in F (the words with a constituent represented by a dark blue vertex in Figure 1):

$$S = G \setminus F. \quad (1)$$

Just as the primary family size measure, the secondary family size measure is log-transformed to remove a strong rightward skew from its distribution.

Formal description of the Rescorla-Wagner equations

Let PRESENT(X,t) denote the presence of cue (letter unigram or letter bigram) or outcome (meaning) X at time t, and ABSENT(X, t) denote its absence at time t. The Rescorla-Wagner equations specify the association strength (or weight) V_i^{t+1} of cue C_i with outcome O at time t + 1 by means of a recurrence relation

$$V_i^{t+1} = V_i^t + \Delta V_i^t. \tag{2}$$

The change in association strength ΔV_i^t is defined as

$$\Delta V_i^t = \begin{cases} 0 & \text{if } \textsc{absent}(C_i, t) \\ \alpha_i \beta_1 \left(\lambda - \sum_{\textsc{present}(C_j,\, t)} V_j \right) & \text{if } \textsc{present}(C_j, t) \ \& \ \textsc{present}(O, t) \\ \alpha_i \beta_2 \left(0 - \sum_{\textsc{present}(C_j,\, t)} V_j \right) & \text{if } \textsc{present}(C_j, t) \ \& \ \textsc{absent}(O, t) \end{cases} \tag{3}$$

Standard settings for the parameters are λ = 1, all α's and β's equal to 0.1.

Items used in Experiment 1 and 2

Between parentheses are the values of the new family size measure obtained from the Family Size Generation Study

Identical cognates: alcohol (5), ark (0), baron (5), camera (5), chaos (2), ego (11), flora (4), globe (5), god (9), horizon (3), hotel (4), lip (6), minister (11), moment (5), norm (14), opera (4), oven (4), psalm (3), shirt (12), sultan (1), tent (1), toilet (8), truck (2), villa (0), volume (4)

English non-identical cognates: admiral (2), advice (8), altar (1), athlete (2), bible (5), camel (3), canal (2), cigar (6), coffee (7), flesh (8), friend (10), honey (9), jewel (7), melon (4), method (7), pill (2), planet (3), prince (4), soup (7), sword (7), tea (25), thief (5), tomato (4), tongue (4), year (9)

English control items: fame (6), throat (6), gun (21), eagle (6), duke (5), widow (3), silk (4), berry (11), fate (9), funeral (4), bench (8), basket (7), lion (5), lad (1), wife (5), noise (6), horse (36), skill (4), donkey (1), torch (1), cellar (3), pigeon (2), bird (26), road (20), animal (5), arrow (2), loss (3), thigh (1), engine (6), window (6), cattle (1), spine (5), carrot (4), tale (6), guilt (6), dusk (1), spider (5), muscle (5), cab (4), wood (37), chest (3), faith (8), wealth (2), sale (13), law (18), frog (7), giant (1), cave (5), peace (16), heaven (4)

Joint analysis of bilingual and monolingual data including *AoA*

Table 1. *Coefficients of the main effects and interaction effects of the GAM. together with the estimate, standard error, t-value, and p-value. The reference values for Identical Cognate are False and Group=Bilinguals. For the non-parametric part of the model, the smooth terms are presented, along with their effective degrees of freedom (edf), reference degrees of freedom (Ref.df), F-value and p-value. Note that the interaction of Trial by Subject is presented by their shrunk factor smooths, and that smooth term for Word represents the by-word random intercepts.*

	Estimate	Std.Error	t-value	p-value
Intercept	-0.7977	0.0866	-9.21	0.0000
Group=Monolinguals	-0.4874	0.0773	-6.31	0.0000
English Frequency	-0.1575	0.0196	-8.03	0.0000
English Primary Family Size	-0.0533	-0.0179	-2.98	0.0029
English Secondary Family size	-0.0009	0.0050	-0.18	0.8568
Identical Cognate=TRUE	-0.1831	0.0351	-5.22	0.0000
Previous RT inverse	0.2180	0.0131	16.62	0.0000
OLD	-0.0645	0.0226	-2.86	0.0044
Group=Monolinguals by OLD	0.0506	0.0215	2.36	0.0183
Group=Monolinguals by Identical Cognate=TRUE	0.0921	0.0244	3.77	0.0002
Group=Monolinguals by English Frequency	0.0758	0.0193	3.65	0.0003
Identical Cognate=TRUE by English Secondary Family Size	0.0287	0.0093	3.10	0.0020
Smooth terms	edf	Ref.df	F	p-value
s (LogAoA): Groupbilinguals	1.00	1.00	1.36	0.2444
s (LogAoA): Groupmonolinguals	5.46	6.35	4.48	0.0001
s (Trial, Subject)	54.53	296.00	2.00	<0.0001
s (Word)	62.00	91.00	2.37	<0.0001

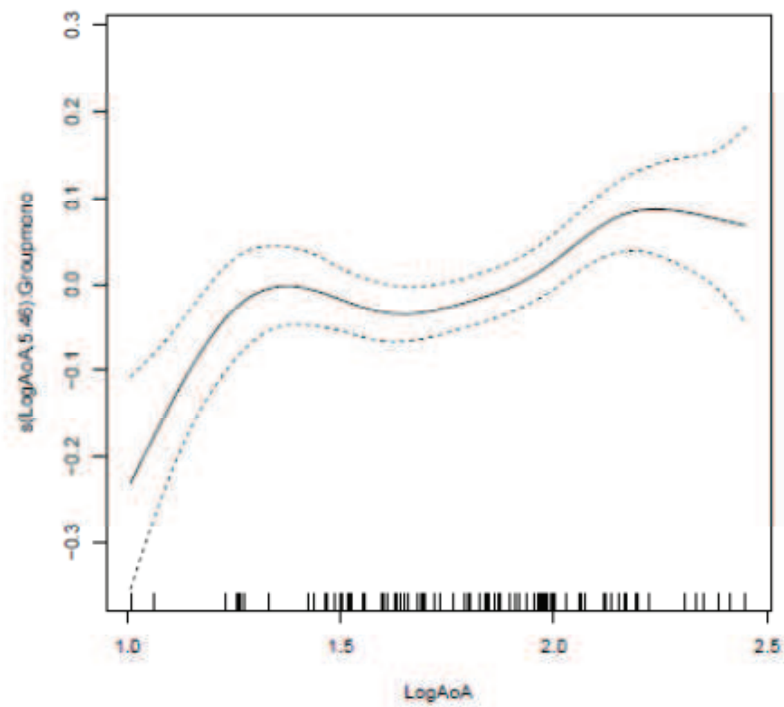


Figure 1. Non-linear effect of Age of Acquisition (log-transformed) in the monolingual group.

Footnotes

Footnote [i]. However, Yarkoni, Balota, and Yap (2008) observed that OLD-20 produced a positive coefficient in their monolingual data. In other words, faster responses were observed when words are more similar to other words. In our bilingual data, the reverse pattern was observed (see also Ferrand et al., 2011, who observed that OLD-20 had little influence on the processing of French monomorphemic words tested in Chronolex). The discrepancy between these results illustrate the inconsistency in findings reported in the literature concerning effects of orthographic similarity (see Ferrand, 2001, for a review). These inconsistencies may be due to several factors, including the distribution of neighbors across the different letter positions in the word and across languages. Furthermore, due to OLD's dependency on a fixed set of 20 words, it may conflate neighborhood density with word frequency (Schepens et al., 2013).

Footnote [ii]. Note, however, that cross-language effects of Dutch orthographic neighborhood size were observed in English lexical decision with Dutch-English bilinguals in a factorial design in which both the English and Dutch neighborhood for English non-cognate words were varied (Van Heuven, Dijkstra, & Grainger, 1998). More research is needed to clarify discrepancies between different cross-language effects.

Footnote [iii]. A generalized additive model (GAM) is an extension of the general linear model that allows the modelling of non-linear relationships between one or more predictors and the dependent variable. It consists of a parametric part that is identical to that of a standard (generalized) linear model, and a non-parametric part that provides functions for modelling non-linear functional relations in two two or higher dimensions. GAMs are especially useful for the modelling of interactions of numerical predictors. Whereas multiplicative interactions in the generalized linear model impose a very specific (and highly restricted) functional form, the so-called tensor product smooths of GAMs make it possible to fit wiggly regression surfaces and hypersurfaces (see Wood, 2006, for further details).

Footnote [iv]. The Hamming distance between two strings of equal length is the number of positions at which the corresponding symbols are different. In other words, it measures the minimum number of substitutions required to change one string into the other.

Footnote [v]. For simulation studies with naive discrimination learning addressing the crucial importance of language-specific distributional properties for understanding cross-linguistic differences in the effects of letter transpositions, see Baayen (2013) in response to Frost (2012).

Table

Table 1. *Item characteristics of the experimental items used in Experiment 1.*

	Identical cognate	Non-identical cognate	Control
Length	5.04	5.28	4.82
English Frequency	2.85	3.07	2.99
English Primary Family Size	2.39	2.62	2.47
English Secondary Family Size	3.68	3.34	4.79
Age of Acquisition	7.39	6.26	6.15
English Bigram Frequency	8.01	8.10	8.02

Table 2. *Coefficients of the main effects and interaction effects of the final model, together with the estimate, standard error, t-value, p-value, and lower and upper 95% confidence intervals in Experiment 1. The reference value for Identical Cognate is False.*

	Estimate	Std. Error	t-value	p-value	Left CI	Right CI
Intercept	-1.5272	0.0868	-17.5868	0.0000	-1.7009	-1.3536
Trial	-0.0149	0.0111	-1.3352	0.1819	-0.0372	0.0074
English Frequency	-0.1006	0.0219	-4.5969	0.0000	-0.1443	-0.0568
English Primary Family Size	-0.0500	-0.0226	-2.2096	0.0272	-0.0952	-0.0047
Identical Cognate=TRUE	-0.0685	0.0451	-1.5197	0.1287	-0.1587	0.0217
English Secondary Family size	-0.0021	0.0062	-0.3378	0.7355	-0.0146	0.0104
Previous RT	0.0777	0.0191	4.0641	0.0000	0.0395	0.1160
Identical Cognate=TRUE by English Secondary Family Size	0.0270	0.0122	2.2087	0.0273	0.0025	0.0514

Table 3. *Coefficients of the main effects and interaction effects of the final model, together with the estimate, standard error, t-value, p-value, and lower and upper 95% confidence intervals in Experiment 2. The reference value for Identical Cognate is False.*

	Estimate	Std. Error	t-value	p-value	Left CI	Right CI
Intercept	-0.9635	0.0934	-10.3130	0.0000	-1.1504	-0.7767
English Frequency	-0.1742	0.0192	-9.0525	0.0000	-0.2126	-0.1357
English Primary Family Size	-0.0540	0.0208	-2.6033	0.0093	-0.0955	-0.0125
Identical Cognate = TRUE	-0.1733	0.0394	-4.3946	0.0000	-0.2522	-0.0944
English Secondary Family Size	0.0004	0.0056	0.0695	0.9446	-0.0108	0.0116
OLD	-0.0709	0.0230	-3.0825	0.0021	-0.1169	-0.0229
PreviousRT	0.1020	0.0168	6.0623	0.0000	0.0683	0.1356
Identical Cognate =TRUE by English Secondary Family Size	0.0255	0.0103	2.4862	0.0130	0.0050	0.0461

Table 4. *Coefficients of the main effects and interaction effects of the model of Simulation Study 1, together with the estimate, standard error, t-value, p-value, and lower and upper 95% confidence intervals.*

	Estimate	Std. Error	t-value	p-value
Intercept	6.7568	0.7474	9.041	<0.0001
English Frequency	-1.0364	0.2321	-4.466	<0.0001
English Primary Family Size	-0.4718	0.2382	-1.980	0.0509
Cognate=Identical	-0.3598	0.5269	-0.683	0.4965
Cognate=Non-identical	-0.0582	0.4989	-0.117	0.9074
English Secondary Family Size	-0.0264	0.0813	-0.324	0.7465
Cognate=Identical by Secondary Family Size	0.1675	0.1355	1.237	0.2196
Cognate=Non-identical by Secondary Family Size	0.0459	0.1401	0.327	0.7441

Table 5. *Coefficients of the main effects and interaction effects of the model of Simulation Study 2a, together with the estimate, standard error, t-value, p-value, and lower and upper 95% confidence intervals.*

	Estimate	Std. Error	t-value	p-value
Intercept	9.9977	0.9256	10.801	<0.0001
English Frequency	-1.3590	0.2219	-6.125	<0.0001
English Primary Family Size	-0.4715	0.2379	-1.982	0.0508
Cognate=Identical	-0.8650	0.4999	-1.730	0.0873
Cognate=Non-identical	0.5626	0.5038	1.117	0.2673
English Secondary Family Size	0.0252	0.0806	0.313	0.7554
OLD	-1.0480	0.2614	-4.010	0.0001
Cognate=Identical by Secondary Family Size	0.0886	0.1271	0.697	0.4879
Cognate=Non-identical by Secondary Family Size	-0.0186	0.1353	-0.138	0.8910

Table 6. *Coefficients of the main effects and interaction effects of the model of Simulation Study 2b, together with the estimate, standard error, t-value, p-value, and lower and upper 95% confidence intervals.*

	Estimate	Std. Error	t-value	p-value
Intercept	6.4330	0.0170	377.556	<0.0001
English Frequency	-0.0229	0.0051	-4.453	<0.0001
English Primary Family Size	-0.0115	0.0056	-2.046	0.0440
Cognate=Identical	-0.0226	0.0118	-1.908	0.0598
Cognate=Non-identical	0.0009	0.0117	0.074	0.9412
English Secondary Family Size	-0.0011	0.0019	-0.598	0.5514
OLD	-0.0152	0.0062	-2.454	0.0162
Cognate=Identical by Secondary Family Size	0.0053	0.0030	1.742	0.0852
Cognate=Non-identical by Secondary Family Size	-0.0001	0.0032	-0.017	0.9866

Table 7. *Generalized additive model predicting identical cognate status for monosyllabic monomorphemic English words.*

A. Parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept	-0.3335	0.3913	-0.8523	0.3940
PC1	-0.1242	0.0372	-3.3413	0.0008
Length	-0.4632	0.0958	-4.8330	<0.0001
Word Category: V	-0.7846	0.1822	-4.3070	<0.0001
Voice: voiceless	0.3991	0.1512	2.6399	0.0083
B. smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
Tensor product smooth Written Frequency and PC2	4.6582	5.4750	42.8571	<0.0001

Table 8. *Generalized additive model fitted to the ELP lexical decision latencies for 2197 monomorphemic monosyllabic nouns and verbs.*

A. Parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept	6.3045	0.1064	59.2480	<0.0001
PC1	0.0026	0.0008	3.2529	0.0012
Voice: voiceless	-0.0087	0.0033	-2.6175	0.0089
Mean Bigram Frequency	0.0085	0.0024	3.4808	0.0005
Noun-Verb Ratio	-0.0021	0.0008	-2.8367	0.0046
Word Category: Verb	-0.0321	0.0920	-0.3486	0.7274
Complex Synsets Count	-0.0061	0.0018	-3.3681	0.0008
Inflectional Entropy	-0.0173	0.0042	-4.1625	<0.0001
Identical Cognate: TRUE	0.0839	0.0867	0.9675	0.3334
B. smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
tensor smooth primary and secondary family size: Nouns	5.5678	6.1607	6.1085	<0.0001
tensor smooth primary and secondary family size: Verbs	5.2049	6.0694	2.4904	0.0206
smooth Written-Spoken Frequency Ratio	2.4997	3.2725	18.8503	<0.0001
Smooth Written Frequency: cognate	4.0177	5.0296	53.8535	<0.0001
Smooth Written Frequency: non-cognate	6.0068	7.0911	91.5963	<0.0001

Figure captions

<u>Figure 1.</u> Activation of primary and secondary family members of the target word *horse*. In Figure 1, a target word, *horse*, is represented by a grey vertex in a directed graph. The directed edge connecting *horse* to *fly* indicates that *horsefly* is an existing compound. The constituents of the compounds in the primary family size of horse are shown with light blue vertices. If activation spreads along the edges of the graph (in both directions, the orientation of the edges only serves to indicate the order of modifier and head), then after having spread into the primary family, it might spread further, leading to the activation of further, semantically more distant, compounds such as *flypaper*, *hairbrush*, and *cocktail*. These more distant compounds are the secondary family members. In Figure 1, the constituents of these secondary family members (when not shared with compounds in the primary family) are represented by dark blue vertices.

<u>Figure 2.</u> Partial effects of the significant predictors on response latencies in English lexical decision Experiment 1 (monolinguals).

<u>Figure 3.</u> Partial effects of the significant predictors on response latencies in English lexical decision Experiment 2 (bilinguals).

<u>Figure 4.</u> Simulated and observed coefficients for the regression models fitted to Experiment 1.

Figure 5. Simulated and observed coefficients for the regression models fitted to Experiment 2, using a single integrated network.

Figure 6. Simulated and observed coefficients for the regression models fitted to Experiment 2, using two separate networks.

Figure 7. Log odds for identical cognate (left panels) as a function of written frequency by PC2. The left panel present the partial effect of the tensor smooth. One standard error confidence regions are denoted by dotted green lines (up) and dashed red lines (down). The right panel presents a contour plot of the same surface, with darker green indicating a lower log odds for cognate status, and colors in pink and white indicating higher log odds for cognate status.

Figure 1

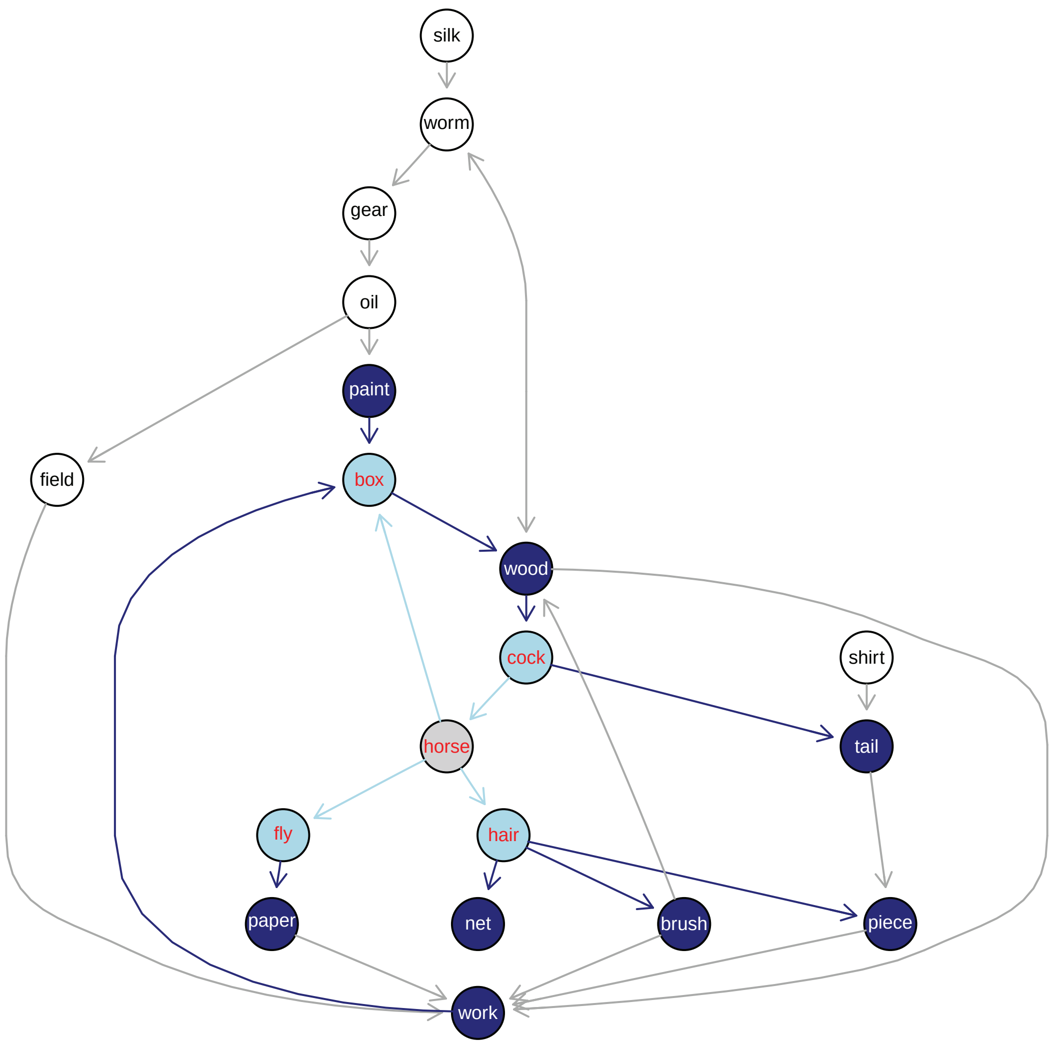

silk
worm
gear
oil
paint
field
box
wood
cock
shirt
horse
tail
fly
hair
paper
net
brush
piece
work

Figure 2

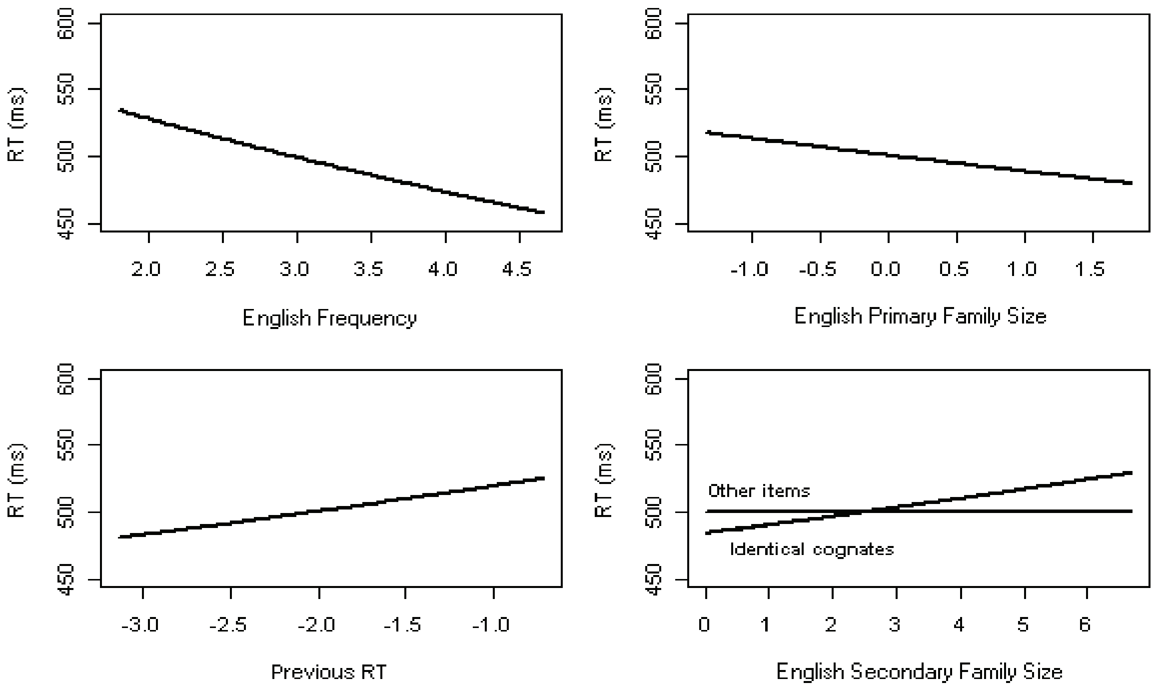

RT (ms)
600
550
500
450
2.0
2.5
3.0
3.5
4.0
4.5
English Frequency
-1.0
-0.5
0.0
0.5
1.0
1.5
English Primary Family Size
-3.0
-2.5
-2.0
-1.5
-1.0
Previous RT
Other items
Identical cognates
0
1
2
3
4
5
6
English Secondary Family Size

Figure 3

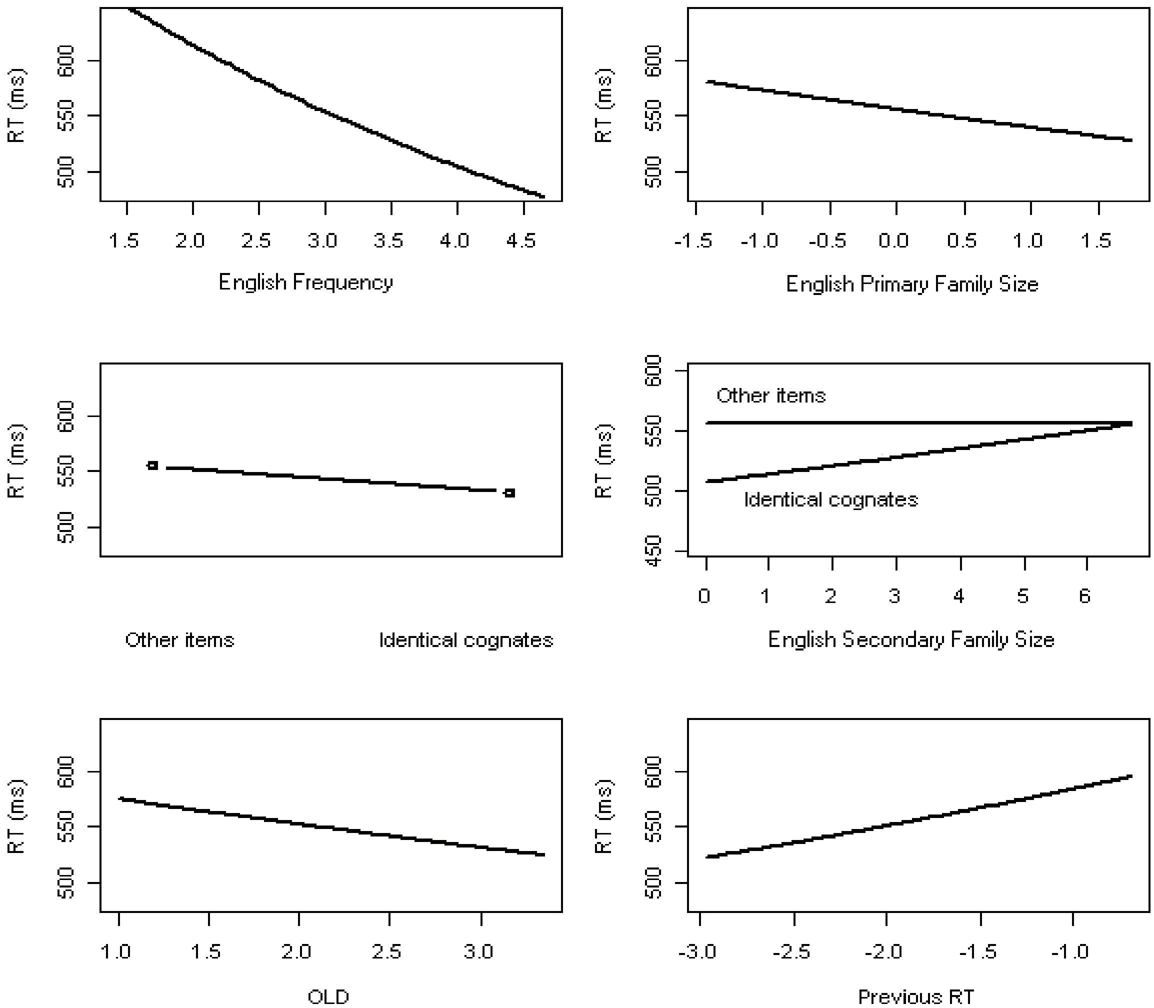

RT (ms)
500
550
600
1.5
2.0
2.5
3.0
3.5
4.0
4.5
English Frequency
RT (ms)
500
550
600
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
English Primary Family Size
RT (ms)
500
550
600
Other items
Identical cognates
RT (ms)
450
500
550
600
Other items
Identical cognates
0
1
2
3
4
5
6
English Secondary Family Size
RT (ms)
500
550
600
1.0
1.5
2.0
2.5
3.0
OLD
RT (ms)
500
550
600
-3.0
-2.5
-2.0
-1.5
-1.0
Previous RT

Figure 4

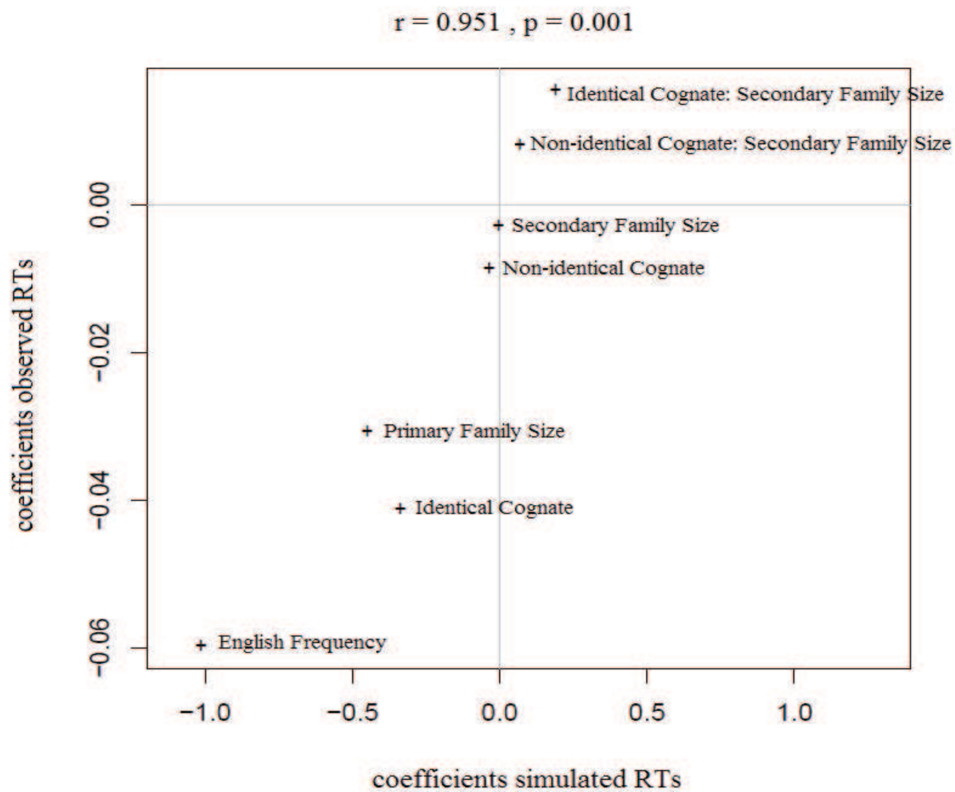

r = 0.951 , p = 0.001
Identical Cognate: Secondary Family Size
Non-identical Cognate: Secondary Family Size
Secondary Family Size
Non-identical Cognate
Primary Family Size
Identical Cognate
English Frequency
0.00
−0.02
−0.04
−0.06
−1.0
−0.5
0.0
0.5
1.0
coefficients observed RTs
coefficients simulated RTs

Figure 5

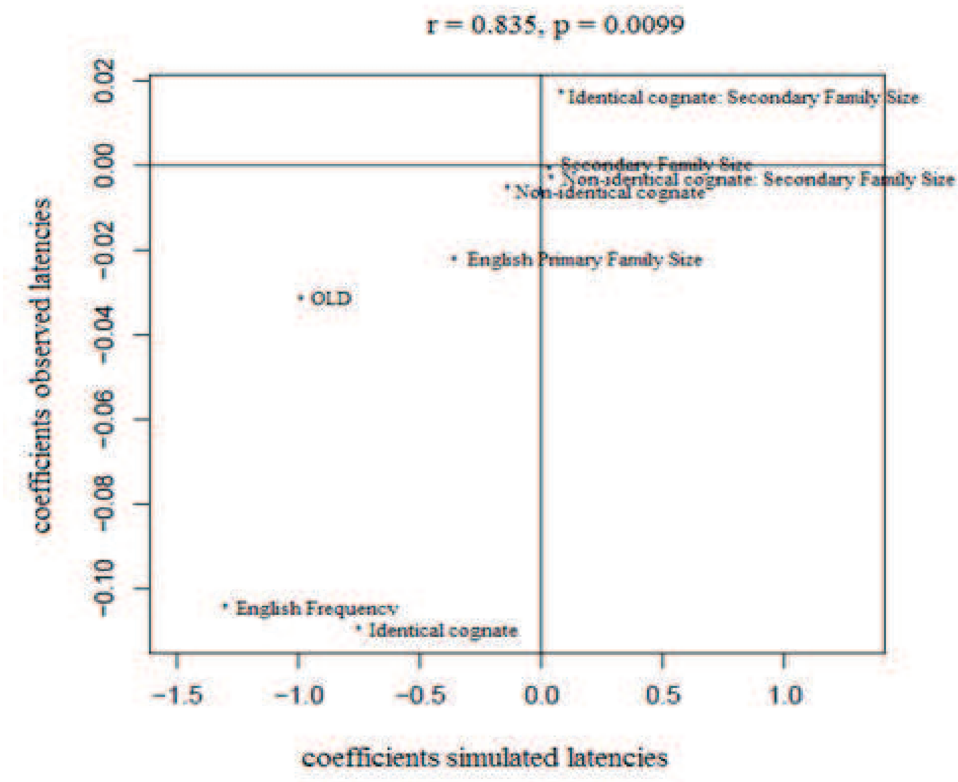
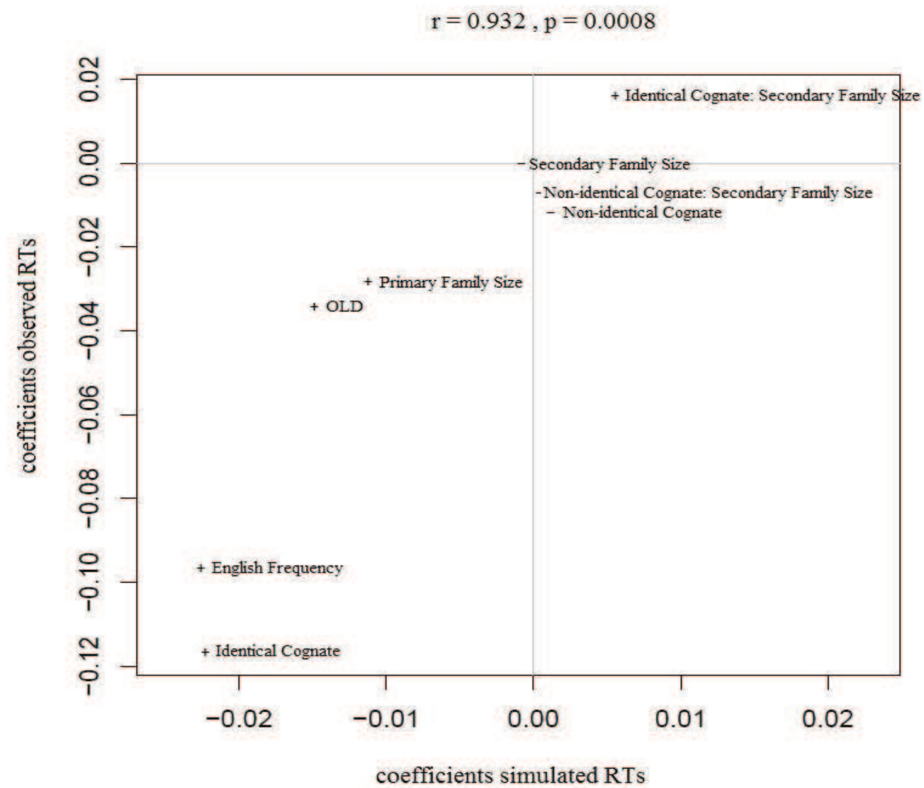

r = 0.835, p = 0.0099
Identical cognate: Secondary Family Size
Secondary Family Size
Non-identical cognate: Secondary Family Size
Non-identical cognate
English Primary Family Size
OLD
English Frequency
Identical cognate
0.02
0.00
−0.02
−0.04
−0.06
−0.08
−0.10
coefficients observed latencies
−1.5
−1.0
−0.5
0.0
0.5
1.0
coefficients simulated latencies

Figure 6



r = 0.932 , p = 0.0008
+ Identical Cognate: Secondary Family Size
− Secondary Family Size
- Non-identical Cognate: Secondary Family Size
− Non-identical Cognate
+ Primary Family Size
+ OLD
+ English Frequency
+ Identical Cognate
coefficients observed RTs
0.02
0.00
−0.02
−0.04
−0.06
−0.08
−0.10
−0.12
−0.02
−0.01
0.00
0.01
0.02
coefficients simulated RTs

Figure 7