Kristoffer Holmquist
ENGK01

Lexical processing of compound words in a L2

A reaction time-based investigation of morphological structure sensitivity in Swedish-English bilinguals

Author: Kristoffer Holmquist

English Studies

Autumn Term 2015

Course: ENGK01

Supervisor: Henrik Gyllstad

Centre for Languages and Literature

Lund University

Kristoffer Holmquist
ENGK01

Abstract

This paper investigates the workings of the mental lexicon in bilinguals and examines how a bilingual speaker may store and compute different morphological structures of nouns. The study investigates the claim that recognising a translation from a compound to another compound should be faster and more accurate than recognising a translation from a compound to a monomorphemic word. The participants of the study were bilingual speakers of Swedish and English (*N* = 16) to whom a translation recognition task was administered consisting of backwards translation from their L2 (English) to their L1 (Swedish), where the English words were compound nouns and the Swedish words were either compound nouns or monomorphemic nouns.

The items selected for the task consisted of 15 compound-to-compound translation equivalents, 15 compound-to-monomorphemic equivalents, and 30 filler pairs that were not translations and only indirectly contributed to the results. The positive items were all semantically transparent nouns, and the compounds all had Noun + Noun structure. All words were tested for length and frequency. The participants were presented with the task in the software DMDX and either accepted or rejected each translation pair as it was displayed. Their reaction time and error rate regarding the various translations were collected and analysed for patterns within the conditions of compound-to-compound and compound-to-monomorphemic translation. After the task, the participants were asked to fill in a frequency-based vocabulary test, self-evaluate their skills in English and complete a short questionnaire pertaining to their age of acquisition and daily use of the language.

According to the findings, compound-to-compound recognition is on average quicker and results in fewer errors than recognition of an English compound to a Swedish monomorphemic word. This could mean that priming in the bilingual lexicon is restricted in terms of morphological structure; compounds activate each other cross-linguistically to a greater degree, whereas, even in proficient bilingual users, monomorphemic translation equivalents are not similarly associated. This reflects what is known as shared representations across language boundaries; a speaker expects translation equivalents to adhere to the same template, which facilitates recognition when this is the case.

Table of contents

1. Introduction

1.1 Introduction

Previous research on lexical processing in bilinguals[1] has focused on the form and meaning of words and translation pairs. Tokowicz (2015) argues that this is an exciting topic as it could tell us much of mental structure in bilinguals, particularly ”how different types of words may be represented in the bilingual mind, or what types of words may be easier or harder to learn/process” (p. 75). Many studies deal with the processing of cognate and non-cognate translations in a speaker’s native language (L1) and second language (L2) and whether there is a difference in processing between translation pairs that are orthographically or phonologically similar compared to those that are not (e.g., Sánchez-Casas & García-Albea, 2005), how concrete and abstract words may differ as translation equivalents (e.g., de Groot, A. M., Dannenburg, L., & van Hell, J. G., 1994), and how speakers react to polysemes with more than one possible translation in a language (Tokowicz & Kroll, 2007).

A type of word that has received less attention in the field of bilingual lexical processing is the compound. There have been attempts to formulate criteria for what actually constitutes a compound, but this is easier said than done as this class of words has, per definition, so few constraints. A more detailed description of compounding will be given in section 1.2; for now, it is enough to say that compounds are words that consist of several lexical morphemes that together form a word with new meaning.

This paper investigates compound processing in the L2. More specifically, it investigates a claim made by Levy, Goral & Obler (2006): that bilingual speakers are sensitive to the morphological structure of translation equivalents in the languages they speak – if the word in question is a bimorphemic compound or a monomorphemic word. The authors report that preliminary evidence suggests that translation equivalents with the same morphological structure (e.g. where both are compounds) are more readily considered equivalents than pairs that are a compound word in one language and a monomorphemic word in the other (Blekher, 2004; Goral et al., in preparation).

Thus, the research question addressed in this essay is: for L1 Swedish learners of English as a L2, is there a processing cost for translation pairs consisting of an English compound and a

[1] There are different definitions of bilingualism; Wei (2000) compares simultaneous bilinguals that grow up in bilingual homes to secondary bilinguals (who are native speakers of only one language and learn their second in another environment). In this essay, the term “bilingual” refers to secondary bilinguals who are native speakers of Swedish and have acquired English later in life.

Swedish monomorphemic word compared to translation pairs consisting of an English compound and a Swedish compound?

The rest of this essay will be structured as follows. First, compounding and the theoretical background of this study will be explained more in-depth. Next, the design of the experiment, participant characteristics and experiment procedure will be clarified. Finally, the results will be presented and discussed. The essay will end with a conclusion section.

1.2 Background

Central to this essay is the notion of compound words. Plag (2003) starts off with the definition that a compound can be described "rather loosely as the combination of two words to form a new word" (p. 133). This is followed by stating that the two assumptions made – that a compound can only consist of two elements, and that the elements are words in themselves – need to be investigated. He emphasises that compounds often consist of more than two elements, providing examples like *power source requirement* and the even more complex *university teaching award committee member*. All compounds, however, are alike in that they have a head – a core that the other elements modify. In English, most compounds are interpreted so that "the left-handed member somehow modifies the right-hand member" (p. 135); a *film society* being a kind of society and a *bookshelf* being a kind of shelf. The head mostly governs the semantic and syntactic properties of the compound. Looking at what kind of elements can be part of compounds, not all are words that appear by themselves; considering *astrophysics* and *biochemistry*, *astro* and *bio* do not occur outside of compounds. They are rather to be considered roots – however, because we have compound words like *teeth marks* and *over-the-fence gossip*, it is clear that not all compound elements are roots. Plag determines that a compound consists of elements that can be either roots, words or even phrases.

Plag regards compounding as "the most productive type of word-formation process in English" (p. 132), but points out that it may also be the most controversial one. Jackendoff (2002) says that the lack of limitations to compounding indicates that this type of word has existed in language since its earliest form. Dressler (2006) attests that compounds exist in virtually all languages, and that different compounds may have vastly different structure – Noun + Noun (*doorknob*), Adjective + Noun (*blueberry*), Verb + Noun (*pickpocket*), and many more variants. In terms of orthography, there are three ways of spelling compounds in English: with a space (roof tile), without a space (rooftile) and with a hyphen (roof-tile). Compounds are of great interest grammatically, but also psycholinguistically. Libben (2006) discusses why they should be studied with such great interest in the latter field. Compounds are phenomena seemingly so central to language that that even other species like chimpanzees can produce novel ones – Libben calls compounding a “universally fundamental word formation process” (p. 2) that is very effective in terms of creating and sharing new meanings that may often be immediately understood. This is because, like sentences are segmentable into words, compounds are segmentable into more or less semantically transparent constituents. Libben advocates the theory that the human brain takes two routes to maximize the chance of understanding a compound: decomposition, where the constituents are separated and analysed individually, and storage of a compound as a whole unit. While a frequent enough compound is likely to be stored as a single unit, this does not happen at the expense of decomposition. As such, the study of compounds occupies a special place in morphology and provides a unique chance to study the workings and structure of the mental lexicon. The idea of decomposition is going to play a vital role in the problem at hand.

While there is a growing body of psycholinguistic research focusing on monolingual speakers, there is still much to do in the field of multilingualism. Jarema (2006) brings up compound processing from a cross-linguistic viewpoint, arguing that studies across language boundaries that compare findings in different systems may yet yield many answers on the topic of Second Language Studies. Similar research has been carried out by different people in different languages, like the role of transparency and priming effects on compounds in English (Libben, 1998; Libben *et al.* 2003) and Dutch (Sandra, 1990; Zwitserlood, 1994); still, studies that focus on multilingual speakers are still scarce in the field – Jarema calls them “surprisingly rare” (p. 67). Comparing results from speakers of different languages may help in arriving at more interesting conclusions, and it is hoped that this will encourage a rise in cross-linguistic studies. This is quite a large gap in the research; there are extensive studies of compound words in single languages, but not much that investigates the relationship between different languages in a speaker’s mind. The bilingual lexicon is an interesting issue as further investigations could tell us much about how lexical processing works in a speaker’s L1 and L2, and what kind of rules and patterns a speaker may attempt to transfer across language boundaries.

Looking at multilingualism psycholinguistically, very intriguing questions arise regarding mental structure: how does the mind handle translation equivalents of compound words, particularly when they have different structure across language boundaries? How are words that have the “same” meaning but are processed differently actually recognised and represented in a bilingual speaker’s lexicon? When we consider compounds words, that by their very nature consist of several morphemes, and compare them to monomorphemic words, we have good reason to think the two categories are going to be represented differently; existing research claims that a bilingual speaker should be less likely to automatically associate a compound in one language with a monomorphemic translation equivalent in another, as the bilingual lexicon assumes that a compound in e.g. a speaker's L2 should also be a compound in their L1. Accepting a translation with an entirely different structure should take longer and potentially result in mistakes. The question whether morphological structure and decomposition affects association of two semantically linked words could tell us much about whether and how the mind expects similarities cross-linguistically.

Levy, Goral & Obler (2006) investigate representation and processing of compounds cross-linguistically. They discuss priming effects, the role of proficiency, and – to this essay, most importantly – mental structure. On the topic of transparency and processing, they discuss the results of Sandra (1990) and Libben *et al.* (2003). These sources describe experiments that arrive at different conclusions. Sandra reports that semantically associated primes shortened response time to transparent compounds, the reason being they were decomposed whereas opaque ones were not. Libben *et al.* (2003), however, found that priming effects influence the processing of some more opaque compounds as well. All things considered, it seems that transparency does affect a speaker's retrieval of lexical items – transparent compounds are subject to decomposition whereas opaque are more likely to be retrieved as a single unit.

This rather naturally leads to the question of words that have different transparency across languages. A compound in one language may look entirely different in another, particularly as there are so many types of compounds with constituents from various word classes. Referring back to Libben (2006), lexical processing of such words is going to be affected by headedness and semantic transparency; a fully transparent (TT) compound – like "blueberry" – where the meaning of each constituent contributes to the meaning of the whole is going to be decomposed, whereas a fully opaque (OO) one like "hogwash" will not. If two words with the same meaning are not similarly decomposed due to difference in transparency, it could be that they will not be processed at the same rate.

Chances are this hypothesis could apply to anything that is not decomposed; this question becomes even more puzzling as we consider words that may be compounds in one language but have no polymorphemic equivalent in the other (or have several equivalents of varying morphological structure). Levy, Goral & Obler ask how, for instance, a French word like "niche" (doghouse) is represented in a bilingual's mental lexicon – whether there is an empty slot, or "doghouse" is retrieved as a unit and considered a direct equivalent by French-English bilinguals. Referring back to an earlier mentioned bilingual girl who called a doghouse “chien-maison”, it is suggested that a word being a compound in a speaker’s native language may lead the speaker to believe that the translation equivalent in their L2 is also a compound. It is then stated that when the equivalent is in fact not a decomposable compound, the bilingual lexicon’s search for one will potentially result in longer processing time and transfer errors (see Figure 1, where the empty slots represent absence of compound equivalents and the speaker needs to search for another word).

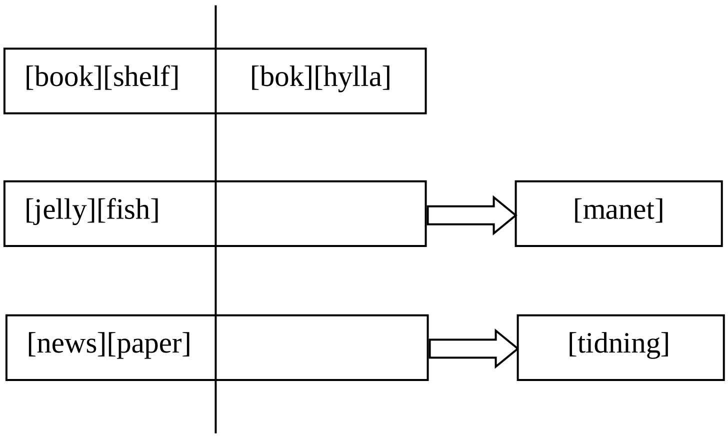

Figure 1: Visualization of structure in the bilingual lexicon. Representation of English compounds and Swedish translation equivalents.

This claim sounds quite plausible in the context, but actually lacks evidence; the authors refer to two sources, “Blekher (2004)” and “Goral *et al.* (in preparation)”, neither of which seems possible to find[2] – nor, to the best of my knowledge, does any other obtainable study deal with this particular question. Therefore, the purpose of this paper is to examine whether the claim that backwards (L2 to L1) translation from a compound to a monomorphemic word will cost processing time and result in more errors holds true for Swedish-English bilinguals. My prediction is that participants will indeed process and recognise compound translations quicker in general, due to compounds in the two languages following the same template; a bilingual speaker is predicted to transfer expectations on morphological structure from one language to another.

[2] Efforts made to retrieve these sources were unsuccessful; Blekher was contacted, but replied that the poster source in question could not be found. Goral et al. has not been successfully located either – presumably, it was never published.

2. Method and materials

2.1 Experiment design

The research question was addressed by conducting a reaction time (RT) experiment. This experiment used a "translation recognition task", where a participant either accepts or rejects a translation. A simple translation task would require the participants to produce their own translation of a word, and this would be a less effective measurement of recognition and likely result in a lower response rate. As the answer I wanted to arrive at was recognition time, a translation recognition task felt more appropriate as it would minimize production on the participants' part. Jiang (2012) discusses de Groot (1992, Experiment 2) and emphasises that a translation recognition task is superior to a simple translation task in that "it helps to reduce missing responses and the feeling of embarrassment on the part of the participants where participants do not know a translation in a regular simple translation task. Second, as no production is required, any effect observed will be that of recognition" (p. 167). Sunderman (2013) says that "simply put, the task measures how quickly and accurately an individual is able to look at two words (presented one after the other) and decide if those two words mean the same thing, or are translation equivalents" (p. 188). According to Sunderman, longer RT and more mistakes in one condition compared to another could indicate the existence of the interference we are looking for.

The software used was DMDX on a laptop computer running Windows. DMDX is a software "designed to precisely time the presentation of text, audio, graphical, and video material and to enable the measurement of reaction times (RTs) to these displays with millisecond accuracy" (Forster & Forster, 2003, p. 116). DMDX is run in a reaction time mode where the participant "makes a binary classification of the input" (p. 122). The software runs on a user-created script created in a text editor and saved in RTF format.

2.2 Item selection

The experiment used 30 positive trials to investigate the possibility of a difference in processing time between two categories of words (see Appendix 1). 15 pairs consisted of bimorphemic compound nouns in English that translated to bimorphemic compound nouns in Swedish (henceforth called C-C+[3]), and the other 15 were bimorphemic compounds in

[3] The plus sign at the end indicate the positive trials and the minus sign the negative trials.

English that translated to monomorphemic words in Swedish (C-M+). All compounds had Noun + Noun structure (like *life* + *guard*). Some of the words I came up with randomly, but many were drawn from a database of 629 compound words compiled by Juhasz, Lai & Woodcock (2015). I attempted to keep the mean frequencies and lengths of the words in the respective conditions as close to each other as possible to try to control for such effects. A comparison of the mean values (one-way ANOVA; see Appendix 5) of the experiment conditions shown in Appendix 1 revealed that for word frequencies, however, there was a difference, F (3, 56) = 9.75, $p < .001$. In terms of relevance for the task at hand, in a post-hoc test, this statistical difference was shown to be located in a comparison of the means of the Swedish compounds and the Swedish monomorphemic words (p = .029), whereas no difference was found between the two groups of English compounds (p = .999). A similar analysis for word length also showed a difference, F (3, 56) = 21.39, $p < .001$. Here, post-hoc pairwise comparisons pointed to a difference between the lengths of the Swedish compounds and the Swedish monomorphemic words ($p < .001$), whereas no difference was found between the two groups of English compounds (p = .091). According to New, Ferrand, Pallier & Brysbaert (2006), word length in the range of 5-8 characters will usually neither facilitate nor inhibit word recognition, and so all Swedish words, compounds and monomorphemic words alike, could presumably fall within the range of 5-8 letters without interfering with RT results. The frequencies of the English words were checked in the Corpus of Contemporary American English (Davis, 2008), and the Swedish words were checked in Korp (Borin, Forsberg & Roxendal, 2012).

The experiment also used 30 filler items in the form of negative trials, with 15 of each category (compound-to-compound, C-C-, and compound-to-monomorphemic, C-M-) to counterbalance the positive sets. These words were not checked for length or frequency. While the filler words were not translations of each other and did not directly contribute in answering the research question, they were imperative to the experiment in preventing participants from understanding the purpose of the task and to keep the probability of a yes or no item at 0.5 throughout the experiment. Jiang (2012) says that critical items can be either positive or negative trials, while Sunderman (2013) goes so far as to say that the negative trials (where the participant is predicted to reject the translation) are the most critical ones in a translation recognition task, but this seems to apply to research where the non-translation items are e.g. morphologically or semantically manipulated and the experiment revolves around the role of distractions. While the critical items in this experiment were in the positive pairs, the procedure also relied on letting the participants both accept and reject translations.

2.3 Participants

An RT experiment is a convenient type of study as it is quite easy to keep manageable. Jiang (2012) gives examples of published RT experiments where no more than twelve to twenty people were tested in total, and he stresses that "RT data do not change significantly after a certain number of participants are tested" (p. 46). According to Jiang, adding another ten participants after ten have already been tested will not significantly alter the results, and thus the sixteen participants in this study should be enough to arrive at a conclusion.

The sixteen participants in this experiment were all native speakers of Swedish and second-language speakers of English, and the majority were undergraduates of English or Linguistics. One was a graduate student, and five were in unrelated fields but still relatively advanced speakers. This was a convenience sample of participants; all were people I knew personally and thought would be interested in participating, and I recruited them by reaching out directly with personal messages. A summary of information about the participants is provided in Table 1.

			Self-report proficiency scores			
	Mean age	Sex (M/F)	Speaking	Listening	Reading	Writing
Participants (*N* = 16)	23.8 (2.5)	56/44%	7.8 (1.1)	8.3 (1)	8.4 (0.8)	7.6 (1.3)
L2 onset	6.5 (1.7)					

Age range = 20-31 years. Use of English: uniform answer was "daily".

Table 1: Participant details. Standard deviation within parentheses.

2.4 Proficiency test

In an attempt to measure the participants' proficiency, I created a short frequency-based vocabulary test with multiple-choice answers to go along with the word recognition task. Rather than administering a large and time-consuming proficiency test to each participant, this type of task can be used as a proxy to determine a speaker's general proficiency as there is empirical evidence that vocabulary scores are highly correlated with measures of reading, writing, listening and speaking (Alderson, 2005). I used Nation & Beglar's (2007) Vocabulary

Size Test as a starting point and randomly selected two words from each of the fourteen frequency bands. The test then consisted of 28 words and measured a vocabulary size up to the 14,000 most common words in English. Each word was shown in a short, non-defining context, followed by four options including one correct definition. Nation (2012) says that the test is designed to measure the skills of both first- and second-language speakers of English, but that it is important to remember that it only measures receptive knowledge and is not a complete measurement of reading skill, nor of how well a speaker may use their vocabulary actively. This type of test is not a perfect way of determining proficiency, but provides a good enough approximation of how advanced a speaker can be expected to be.

2.5 Procedure

The participants were tested individually and instructed in English. First, they were asked to engage in the translation recognition task. The experiment was based on backwards translation from the participants' L2 to their L1 to compare the importance of lexical and conceptual links in the mental lexicon. English and Swedish words were shown in pairs; the sequence of each trial is visualized in Figure 2.

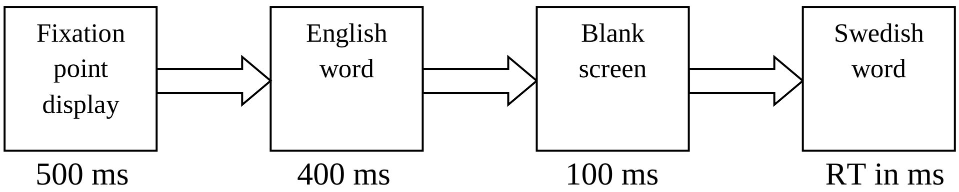

Figure 2: Sequence of presentation in the translation recognition task. The fixation point consisted of eight asterisks displayed in the same location as the subsequent words.

An English compound was shown for 400 milliseconds, and after a 100-millisecond break a Swedish compound or monomorphemic word was shown for up to 4000 milliseconds after which it timed out. A similar sequence is exemplified in Jiang (2012), who refers to Talamas, Kroll & Dufour (1999) (p. 168). Other studies mentioned in the text only displayed the first stimulus for 240 milliseconds while others used up to 500 milliseconds. One reason to use a shorter duration may be to hinder participants from generating a translation too early, but Jiang stresses that how "procedural differences affect participants' performance is yet to be determined" (p. 168). Allowing for up to 4000 milliseconds until an answer was required was a strategy to try to increase the number of responses.

The participants were instructed to press the "Yes" button if the words were translation equivalents and "No" if they were not, and to answer as quickly and accurately as possible. The 60 word pairs were divided into 5 groups of 12, with 3 words from each of the two relevant categories and 3 words from each of the two filler categories in each group. The groups appeared in random order, and the order of the items was also randomized within each group. This approach minimized the risk of too many positive or negative trials in sequence. The participants were shown and reacted to 5 translation words for practice before the actual experiment commenced. The task was completed without pauses, as it was deemed short enough not to fatigue the participants.

After the computer task, the participants completed the frequency-based vocabulary test in an attempt to approximate their L2 proficiency, and answered a questionnaire noting their age, gender, vision, native language, approximate age of L2 onset and how often they use English in their daily life. The questionnaire concluded with a self-evaluation task where they rated their own skills in speaking, listening, reading and writing on a scale from 1-10 where 1 corresponded to no proficiency and 10 to a native speaker's level. The proficiency testing seemed more appropriate to administer after the experiment itself, not least to avoid the possibility of participants feeling fatigued or discouraged when faced with the actual RT study. The whole experiment took approximately 15-20 minutes per participant, with the translation recognition task only taking a few minutes and the vocabulary size test usually requiring 5-10 minutes to complete. The questionnaire and self-evaluation was very quick; no participants seemed to hesitate how to rank their own skills.

2.6 Data analysis and proficiency assessment

All response times to the positive items were analysed while the negative pairs were ignored. I calculated the mean reaction time to translation pairs the participants had got right, noting number of errors but obviously leaving mistakes out of the average score. The prediction was that mean RT to the compound-to-compound condition would be shorter than the compound-to-monomorphemic condition. The error rate – erroneously saying "No" to a correct translation – was also predicted to be higher when faced with a monomorphemic equivalent in Swedish.

The 28-item vocabulary test was corrected, with one point awarded for each answer, and the number of accurate answers on each test multiplied by 500 to get an idea of the participant's approximate vocabulary size. Checking and showing the proficiency of the participants was vital to the experiment, as one could expect different reaction times from speakers with too different L2 proficiency and the aim was to recruit a homogeneously advanced L2 proficiency group of participants.

3. Results

3.1 L2 proficiency test

All 16 participants' data were analysed and included in the paper. The vocabulary proficiency test was administered just after the translation task itself. According to the results, the students tested had vocabularies ranging in size from 7,500 to 14,000 words. Discounting the two participants who scored 7,500 and 8,000, the third lowest score was 11,000 words. All participants were thus intermediate or advanced users, and should be proficient enough to recognise and parse the experiment items without problems. The mean score and standard deviations (SD) are displayed in Table 2.

	Mean	SD	Extrapolated mean vocab size (SD)
Participants (*N* = 16)	22.25	3.92	11,125 (1,962)

Table 2: Vocabulary Size Test results.

3.2 Reaction time results

The average reaction time to compound-to-compound translation among the sixteen students was 728 milliseconds while compound-to-monomorphemic translation took 828 milliseconds on average. However, one participant was on average only 2.5 milliseconds quicker with compound-to-compound translation while another was over 400 milliseconds quicker. Most participants ended up with an average categorical RT difference of 50-250 milliseconds, with translation from compounds to compounds consistently being quicker. Only one participant (L12) had a quicker average response time to compound-to-monomorphemic words than compound-to-compound translation, but this person also made the most errors when translating to monomorphemic words: 5 (33%) out of 15, compared to only one error (7%) for compounds. The participants' mean reaction times and error rates are presented in Table 3. The results are visualized in Figures 3 and 4.

	Item conditions			
	C-C+	C-M+	C-C-	C-M-
Participants (*N* = 16)	728(182)	828(192)	867(190)	896(249)
	[3.8%]	[13.3%]	[2.1%]	[0.4%]

Table 3: Mean response time in milliseconds, standard deviations (parentheses) and error rates [square brackets] for the four experimental conditions.

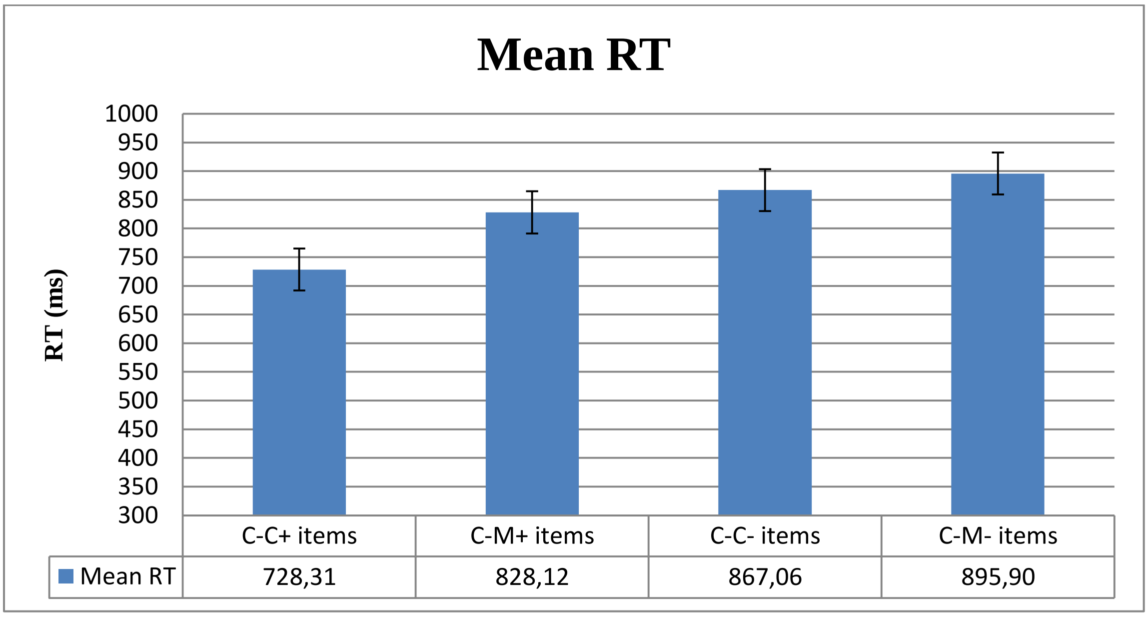

Figure 3: Visualization of RT results to the four conditions.

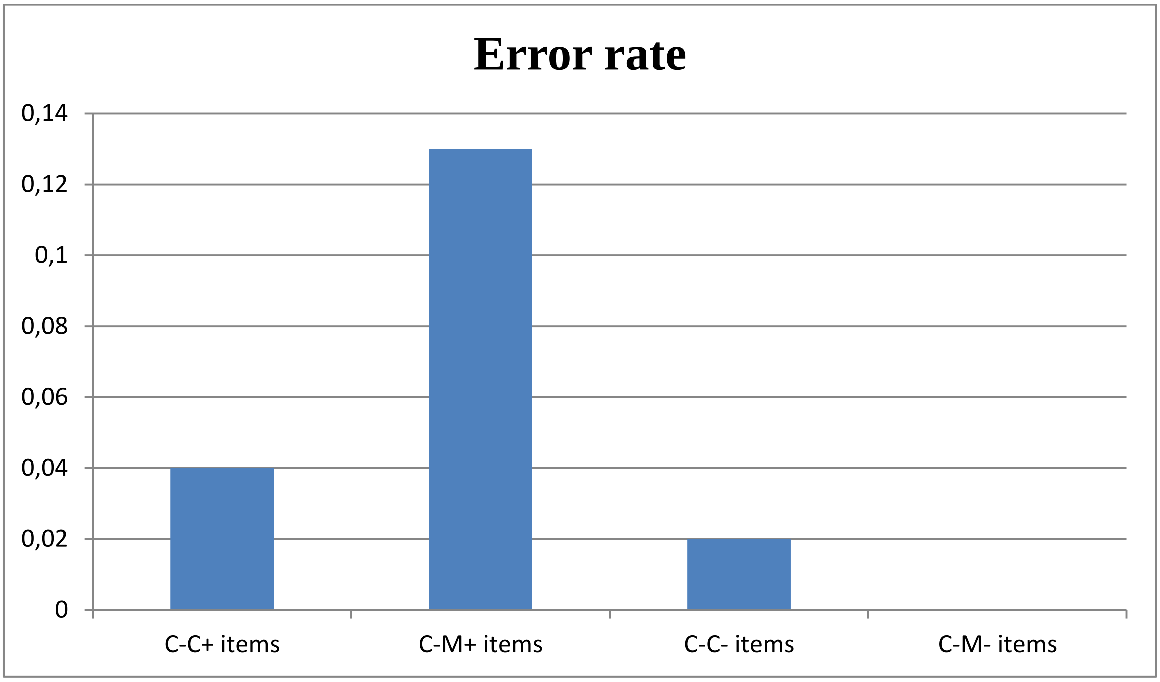

Figure 4: Visualization of error rates. Y axis symbolizes the proportion of errors made to each condition.

A repeated measures ANOVA (see Appendix 5) was used to analyse the data from the translation recognition experiment (Sunderman, 2014, p. 201). The assumption of sphericity was not violated. The results show that reaction times were significantly affected by the type of translation pair, F (3, 45) = 5.18, p < .01, η^2 = .26. Subsequent pairwise comparisons revealed that the mean reaction times between the positive compound to compound condition (C-C+) was significantly different from all of the other three conditions (at p = .011, p = .006, and p = .005, respectively), but for no other pairwise comparisons were statistically significant differences observed.

A few of the participants made significantly more errors than the others, raising the average considerably. Some trials had particularly high error rates – 14 out of the 16 participants got one particular positive trial wrong: C-M+ pair *skullcap* and *kalott* (see item 213 in Appendix 1).

It took significantly longer for the participants to reject negative pairs than to accept positive ones. In the case of compound-to-compound pairs (C-C-), the process was however slightly quicker – the English word was decomposed and recognised, but the Swedish constituents did not match the English ones, and so the translations were rejected. Rejecting monomorphemic non-translations (C-M-) took even longer (896 milliseconds on average).

4. Discussion

The purpose of this study was to investigate the validity of the claim that mental decomposition facilitates compound recognition cross-linguistically while monomorphemic translation equivalents take longer to accept. The results of this experiment with Swedish-English bilinguals point rather definitively in one direction, even though there are differences between some of the participants. The results in general support the claim that translation from a compound to another compound is quicker and more accurate than translation from a compound to a monomorphemic word – and this despite the fact that, as previously mentioned, the Swedish monomorphemic words were shorter and significantly more frequent. C-C+ translation clearly stands out as significantly faster than the other critical condition C-M+. This takes us back to the notion of online decomposition; a result like this could be predicted from the start, based on how we understand lexical processing. Li, Jiang & Gor (2015) discuss the “sublexical” model, which suggests that individual morphemes in complex words are activated before the representation of the word as a whole. In their study, Chinese-English bilingual participants turned out to be “sensitive to the morphological structure of compounding and to decompose compounds rapidly and automatically at the initial stage of visual processing” (p. 15). Based on this, it seems likely that this would also apply to Swedish-English bilinguals; *fire* and *fly* would quickly be recognised as translations of *eld* and *fluga*, which would in turn facilitate recognition of the whole: *eldfluga*. However, when they encounter a compound like *newspaper*, decomposition provides them with *news-nyheter* and *paper-papper,* which cannot be mapped onto the expected template. Instead, they have to search for the monomorphemic equivalent: *tidning*. This is what costs processing time in the RT results. Looking at a compound word like *firefly* and its Swedish translation equivalent *eldfluga* (which constituted a positive translation pair in the experiment), however, we know that when the English word is first displayed, the participant is going to mentally decompose it and interpret the meaning of its constituents. This activates the morphemes’ translation equivalents in Swedish and facilitates recognition of the Swedish compound immediately afterwards. The RT to *firefly–eldfluga* was on average very short, as were the RT to other C-C+ pairs. A monomorphemic translation equivalent is not similarly matched with the morphemes of the compound, and neither do the negative trials with non-translations undergo this process.

The filler items (C-C- and C-M-) were not controlled for length and frequency, and thus the results are not as informative as those of the positive pairs. On average, participants took slightly longer to reject pairs than accept them. With the latter condition, the process may have been similar to positive pairs with a monomorphemic translation equivalent – considering the negative pair *rooftile* and *hallon*, the speaker almost certainly gave “hallon” a chance on the off chance that it actually was a monomorphemic translation, but eventually realised that it was not semantically related in any way.

The somewhat unevenly distributed results also beg the question whether looking at average reaction times is a useful measurement as the participants produce such different results. It may be that the few participants who recognise monomorphemic translations quicker or almost equally quick are exceptions to the rule, but this divergence is still puzzling. Out of the 16 participants, 3 displayed reaction times were very close to each other, or were actually slower to accept compounds.

There could, of course, be a number of factors that influence this kind of task. One could lie in the items selected for the task; referring back to New, Ferrand, Pallier & Brysbaert (2006), word length should not affect recognition as long as it falls within 5-8 characters, which all items did. If anything, the fact that the monomorphemic translation equivalents in Swedish were on average shorter than the Swedish compounds could possibly have facilitated recognition and shortened reaction time to the monomorphemic words. However, Levy, Goral & Obler’s (2006) prediction holds true: compound-to-monomorphemic translation takes slightly longer, and if word length does indeed affect the result in the other direction, it is not significant enough to counter the result we would expect based on what we know about lexical processing. It is highly significant that the results at hand were achieved despite of the monomorphemic words being shorter and more frequent.

All compound words were also semantically transparent, so that the meaning of the whole should have been easy to interpret based on the meaning of the constituents, and the vocabulary proficiency test indicated that no participants should have any problems with the words. Using fully transparent (TT) compounds may have led to a higher degree of decomposition than if the experiment had used opaque (OO) or only partially transparent ones (TO, OT). However, the issue could be one of frequency, tying into the question of the sample of participants. It is possible that a certain participant simply has not encountered some of the words used in the experiment as regularly as he or she has some other words. Even if the items are checked in corpora and word frequencies kept close to each other statistically, such a variable is difficult to control for in relation to every participant. This is illustrated by the translation pair that 14 out of 16 participants erroneously rejected; *skullcap–kalott* is not terribly infrequent when you look at statistics (see item 213 in Appendix 1), but still probably not a word that undergraduate students use very often. Frequency also affects how early and in what way a certain word may have been learnt; a high-frequency word like "backpack" (which made up one half of a positive compound-to-compound pair in the experiment) is likely to have been taught to a Swedish learner of English quite early, and in the early years of school word lists are a common way of learning new vocabulary. This means that practice with word lists and a close to "automatic" association of two words may well affect RT more than morphological structure does – as soon as the English word is displayed, the participant has an ingrained understanding of the "correct" translation and so reacts very quickly.

Another issue influencing RT results could possibly be proficiency in other languages than Swedish and English. 8-9 out of the 15 Swedish monomorphemic words used in the positive trials are loanwords that have possibly entered Swedish directly from French and could be considered to have very similar French equivalents. It is not strange that these have been considered monomorphemic in Swedish since French is further removed from Swedish morphologically than Germanic languages are, but does their monomorphemic status affect their processing. Post-experiment interviews with the participants revealed that some of those who on average reacted equally quick or even quicker to monomorphemic translations were also more proficient in French than the others, having studied it for at least 3 years. However, this did not apply to everyone with unexpected RTs, and there are currently too few examples to conclude that there could be a trend. It is, in a way, not at all new or surprising that cognates in other languages known the participants may facilitate word recognition – a factor I tried to eliminate in the experiment only with regard to English and Swedish words – but it is something to take into consideration when conducting this type of study. Future experiments of this kind could possibly consider what type of influence a speaker's L3 might exert over L1 and L2 processing.

It is important to remember that this experiment was conducted with a convenience sample of participants – local people I personally knew and could reach out to with relative ease. Based on Jiang (2012), we can assume that a sample of 16 is a good number for a RT study, but it still does not reflect a larger population. The participants were all between 20 and 31 years of age, and all were experienced university-level students with intermediate to advanced English vocabularies. Even though the vocabulary proficiency test could be administered and used for analysis of anyone, it is difficult to predict how a control group with different participant ages and educational levels might have performed in this experiment (for instance, it is possible that older native Swedish participants with large English vocabularies might have been more familiar with words like *kalott*, and this would have affected the results).

The results may also have been affected by how the participants approached the task, from a viewpoint relating to semantic transparency. Some participants had never been part of a RT experiment before, took the time aspect very seriously and asked questions about the procedure before we began. They were told that there were no "trick" questions with regard to whether a pair was positive or negative, and that it was going to be rather clear if two words were translations of each other. Since the particularly inquisitive participants were acutely aware of their reactions being timed (which was only implied in the instructions), it could be that structurally similar words were accepted quicker without being fully recognised. For instance, following *firefly*, it is possible that a participant would not take enough time to recognise the whole of *eldfluga*, but click "Yes" as soon as they registered *eld* (*fire*). If this is the case, it is an understandable strategy to minimize reaction time since the negative compound-to-compound pairs were very obvious and easy to recognise, like *fairytale-kaffekopp* and *sketchbook-glädjerus*. This issue could possibly be resolved by the use of eye-tracking in an experiment with similar design and purpose, to determine whether the whole compound is actually recognised or if the participant jumps to conclusions early. At the other end of the spectrum, some participants had a different approach entirely and tried to multitask and start discussing the translations while the experiment was still in progress. There were also some that, according to themselves, took producing a large number of correct responses very seriously and preferred to hesitate and take their time with each trial (scoring a higher mean RT) to minimize the number of errors made.

5. Conclusions

The initial question in this paper concerned whether the difference between mono- and bimorphemic word structure plays a role in recognising translation equivalents in your L1 and L2. This was addressed by conducting a reaction time study with 16 participants that were all undergraduate or graduate students, and L1 speakers of Swedish and L2 speakers of English. The claim in Levy, Goral & Obler (2006), that translation between compounds should be quicker than between a compound and a monomorphemic word and that the latter could result in more errors, seems to hold true for the Swedish-English bilinguals that have been tested.

This seems to mean that the sublexical model that Li, Jiang & Gor (2015) has found support for holds true for bilingual speakers of Swedish and English; the approach that facilitates recognition – mapping decomposed morphemes onto a template directly – fails, as there is no translation equivalent with the same morphological structure. To reach a satisfying result as to whether the difference in RT between L1 compound and monomorphemic translations of L2 compounds sheds light on the structure of a bilingual speaker's lexicon in general, this type of study could be conducted again with L1 speakers of a different language with similar typology. As it stands now, we can certainly say that this seems to be the case for Swedish-English bilinguals.

References

Alderson, J. C. (2005). *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment.* London: Continuum.

Borin, L., Forsberg, M., & Roxendal, J. (2012). *Korp – the corpus infrastructure of Språkbanken. Proceedings of LREC 2012.* Istanbul: ELRA, 474–478.

Davies, M. (2008-). *The Corpus of Contemporary American English: 520 million words, 1990-present.* Available online at http://corpus.byu.edu/coca/.

De Groot, A. M. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 1001–1018.

De Groot, A. M., Dannenburg, L., & van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of memory and language*, *33*(5), 600–629.

Dressler, W. U. (2006). Compound Types. In G. Libben & G. Jarema (Eds.), *The Representation and processing of compound words* (pp. 23–44). Oxford: Oxford University Press.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.

Forster, K. I., Forster J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods*, *Instruments, & Computers*, *35*(1), 116–124.

Jackendoff, R. (2002). *Foundations of Language.* Oxford: Oxford University Press.

Jarema, G. (2006). Compound Representation and Processing: A cross-language perspective. In G. Libben & G. Jarema (Eds.), *The Representation and processing of compound words.* (pp. 45–70). Oxford: Oxford University Press.

Jiang, N. (2012). *Conducting Reaction Time Research in Second Language Studies.* New York: Routledge.

Juhasz, B. J., Lai, Y., & Woodcock, M. L. (2015). A database of 629 English compound words: ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods, 47*(4), 1004–1019.

Levy, E.S., Goral, M., & Obler, L.K. (2006). Doghouse/chien-maison/niche: Compounds in bilinguals. In G. Libben & G. Jarema (Eds.), *The Representation and processing of compound words.* (pp. 125–144). Oxford: Oxford University Press.

Li, M., Jiang, N., & Gor, K. (2015). L1 and L2 processing of compound words: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition,* Available on CJO 2015 doi:10.1017/S1366728915000681

Libben, G. (1998). Semantic transparency in the processing of compounds: consequences for representation, processing, and impairment. *Brain Lang. 61*, 30–44.

Libben, G. (2006). Why study compound processing? An overview of the issues. In G. Libben & G. Jarema (Eds.), *The Representation and processing of compound words* (pp. 1–22). Oxford: Oxford University Press.

Libben G., Gibson M., Yoon Y. B., Sandra D. (2003). Compound fracture: the role of semantic transparency and morphological headedness. *Brain Lang. 84*, 50–64.

Lowie, W., & Seton, B. (2013). *Essential statistics for applied linguistics.* Basingstoke: Palgrave Macmillan.

Nation, I.S.P. (2012). The vocabulary size test. Retrieved from: http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Vocabulary-Size-Test-information-and-specifications.pdf

Nation, I.S.P. & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition. New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review, 13*(1), 45–52.

Plag, I. (2003). *Word-Formation in English.* Cambridge: Cambridge University Press.

Sánchez-Casas, R., & Garcıa-Albea, J. E. (2005). The representation of cognate and noncognate words in bilingual memory. In J. F. Kroll and A.M.B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*, (pp. 226–250). Oxford: Oxford University Press.

Sandra, D. (1990). On the representation and processing of compounds words: Automatic access to constituent morphemes does not occur. *Quarterly Journal of Experimental Psychology, 42A*, 529–567.

Sunderman, G. L. (2013). Translation Recognition Tasks. In Jill Jegerski, & Bill Van Patten (Eds.), *Research Methods in Second Language Psycholinguistics* (pp. 185–211). New York: Routledge.

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second language vocabulary. *Bilingualism: Language and Cognition, 2,* 45–58.

Tokowicz, N. (2015). *Lexical processing and second language acquisition.* London: Routledge.

Tokowicz, N., & Kroll, J. F. (2007). Number of meanings and concreteness: Consequences of ambiguity within and across languages. *Language and Cognitive Processes*, *22*(5), 727–779.

Wei, L. (2000). *The bilingualism reader.* London: Routledge.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes, 9*(3), 341–368.

Appendix 1 – Experiment items

Positive trials

Item No.	**C-C+**	**Length**	**Citation form freq. (SD)**	**Log-normalized frequency**	**Swedish translation**	**Length**	**Freq. (SD)**	**Log-normalized frequency**
101	backyard	8	6566	8.79	bakgård	7	3658	8.20
102	bagpipe	7	124	4.82	säckpipa	8	843	6.74
103	daylight	8	3958	8.28	dagsljus	8	5529	8.62
104	eyelid	6	465	6.14	ögonlock	8	9046	9.11
105	firefly	7	437	6.08	eldfluga	8	272	5.61
106	hourglass	9	372	5.92	timglas	7	853	6.75
107	lifeguard	9	600	6.39	livvakt	7	5815	8.67
108	peacetime	9	710	6.57	fredstid	8	1260	7.14
109	backpack	8	3091	8.04	ryggsäck	8	10607	9.27
110	raincoat	8	657	6.49	regnrock	8	578	6.36
111	foothold	8	774	6.65	fotfäste	8	3005	8.01
112	seaweed	7	968	6.88	sjögräs	7	959	6.87
113	snowball	8	809	6.69	snöboll	7	1972	7.59
114	stairwell	9	1388	7.24	trapphus	8	6578	8.79
115	tombstone	9	633	6.45	gravsten	8	3797	8.24
	Mean	8	1436.8 (1707.65)	6.76		7.67	3651.47 (3131.19)	7.73
	Median	8				8		
	Mode	8				8		
	Range	7-11				7-10		

Item No.	C-M+	Length	Citation form freq. (SD)	Log-normalized frequency	Swe translation	Length	Freq. (SD)	Log-normalized frequency
201	armchair	8	1472	7.29	fåtölj	6	9660	9.18
202	cloakroom	9	106	4.66	garderob	8	45518	10.73
203	drugstore	9	1411	7.25	apotek	6	33824	10.43
204	hairstyle	9	441	6.09	frisyr	6	25698	10.15
205	handgun	7	1522	7.33	pistol	6	15785	9.67
206	doorman	7	763	6.64	portier	7	397	5.98
207	jellyfish	9	532	6.28	manet	5	2024	7.61
208	sidewalk	8	6526	8.78	trottoar	8	9169	9.12
209	paintbrush	10	438	6.08	pensel	6	8280	9.02
210	storeroom	9	420	6.04	förråd	6	13785	9.53
211	pillowcase	10	345	5.84	örngott	7	1564	7.36
212	newspaper	9	26527	10.19	tidning	7	297851	12.6
213	skullcap	8	146	4.98	kalott	6	239	5.48
214	wallpaper	9	1782	7.49	tapet	5	20877	9.95
215	waterfront	10	2171	7.68	strand	6	71075	11.17
	Mean	8.73	2973.47 (6478.63)	6.84	37049.63 (72194.51)	6.33		
	Median	9				6		
	Mode	9				6		
	Range	7-10				5-8		

Negative trials (filler items)

C-C-	Length	Subsequent Swe word	Length	C-M-	Length	Subsequent Swe word	Length
Fruitcake	9	Takterrass	10	Rooftile	8	Hallon	6
Crossroad	9	Fiskpinne	9	Windmill	8	Kruka	5
Glovebox	8	Skogsrå	7	Frostbite	9	Kedja	5
Flowerpot	9	Golvplanka	10	Buttonhole	10	Toffel	6
Sketchbook	10	Glädjerus	9	Cannonball	10	Drake	5
Shipwreck	9	Vakthund	8	Dockyard	8	Tjäle	5
Bookmark	8	Guldgruva	9	Filmstrip	9	Väska	5
Bathroom	8	Nallebjörn	10	Framework	9	Brand	5
Spearhead	9	Handduk	7	Claypole	8	Fodral	6
Molehill	8	Datorspel	9	Oatmeal	7	Kummel	6
Storytime	9	Nyckelring	10	Leapfrog	8	Hövding	7
Pawnshop	8	Skinnsoffa	10	Hailstone	9	Affisch	7
Broomstick	10	Slipsknut	9	Pillbox	7	Smultron	8
Fairytale	9	Kaffekopp	9	Ponytail	8	Hjortron	8
Firewood	8	Leksak	6	Flatmate	8	Mössa	5

Appendix 2 – L2 proficiency test (shortened vocabulary test)

First 1000:

SHOE: Where is your shoe?

a. the person who looks after you

b. the thing you keep your money in

c. the thing you use for writing

d. the thing you wear on your foot

POOR: We are poor.

a. have no money

b. feel happy

c. are very interested

d. do not like to work hard

Second 1000:

UPSET: I am upset.

a. tired

b. famous

c. rich

d. unhappy

PATIENCE: He has no patience.

a. will not wait happily

b. has no free time

c. has no faith

d. does not know what is fair

Third 1000:

RESTORE: It has been restored.

a. said again

b. given to a different person

c. given a lower price

d. made like new again

DASH: They dashed over it.

a. moved quickly

b. moved slowly

c. fought

d. looked quickly

Fourth 1000:

CANDID: Please be candid.

a. be careful

b. show sympathy

c. show fairness to both sides

d. say what you really think

REMEDY: We found a good remedy.

a. way to fix a problem

b. place to eat in public

c. way to prepare food

d. rule about numbers

Fifth 1000:

HAUNT: The house is haunted.

a. full of ornaments

b. rented

c. empty

d. full of ghosts

FRACTURE: They found a fracture.

a. break

b. small piece

c. short coat

d. rare jewel

Sixth 1000:

THRESHOLD: They raised the threshold.

a. flag

b. point or line where something changes

c. roof inside a building

d. cost of borrowing money

MALIGN: His malign influence is still felt.

a. evil

b. good

c. very important

d. secret

Seventh 1000:

QUILT: They made a quilt.

a. statement about who should get their property when they die

b. firm agreement

c. thick warm cover for a bed

d. feather pen

BRISTLE: The bristles are too hard.

a. questions

b. short stiff hairs

c. folding beds

d. bottoms of the shoes

Eighth 1000:

MARROW: This is the marrow.

a. symbol that brings good luck to a team

b. Soft centre of a bone

c. control for guiding a plane

d. increase in salary

AUTHENTIC: It is authentic.

a. real

b. very noisy

c. Old

d. Like a desert

Ninth 1000:

WHIM: He had lots of whims.

a. old gold coins

b. female horses

c. strange ideas with no motive

d. sore red lumps

REGENT: They chose a regent.

a. an irresponsible person

b. a person to run a meeting for a time

c. a ruler acting in place of the king

d. a person to represent them

Tenth 1000:

PEASANTRY: He did a lot for the peasantry.

a. local people

b. place of worship

c. businessmen's club

d. poor farmers

UPBEAT: I'm feeling really upbeat about it.

a. upset

b. good

c. hurt

d. confused

Eleventh 1000:

PALLOR: His pallor caused them concern.

a. his unusually high temperature

b. his lack of interest in anything

c. his group of friends

d. the paleness of his skin

APERITIF: She had an aperitif.

a. a long chair for lying on with just one place to rest an arm

b. a private singing teacher

c. a large hat with tall feathers

d. a drink taken before a meal

Twelfth 1000:

SPLEEN: His spleen was damaged.

a. knee bone

b. organ found near the stomach

c. pipe taking waste water from a house

d. respect for himself

IMPALE: He nearly got impaled.

a. charged with a serious offence

b. put in prison

c. stuck through with a sharp instrument

d. involved in a dispute

Thirteenth 1000:

UBIQUITOUS: Many weeds are ubiquitous.

a. are difficult to get rid of

b. have long, strong roots

c. are found in most countries

d. die away in the winter

DIDACTIC: The story is very didactic.

a. tries hard to teach something

b. is very difficult to believe

c. deals with exciting actions

d. is written in a way which makes the reader unsure of the meaning

Fourteenth 1000:

AUGUR: It augured well.

a. promised good things for the future

b. agreed well with what was expected

c. had a colour that looked good with something else

d. rang with a clear, beautiful sound

LIMPID: He looked into her limpid eyes.

a. clear

b. tearful

c. deep brown

d. beautiful

Appendix 3 – Participant questionnaire

Participant code:

Age: ____

Gender: ______

Vision:

Native language:

First encounter with English:

Use of English:

Here you will be asked to do a self-assessment of your English proficiency. On a scale of 1-10, state your own English proficiency in the four language skills by putting a cross in the corresponding box.

Skills	none 1	2	3	4	5	6	7	8	9	native-like 10
Speaking										
Listening										
Reading										
Writing										

Appendix 4 – Individual RT results and error rates for C-C+ and C-M+ items

Participant	C-C+ Mean RT (ms)	Errors	C-M+ Mean RT (ms)	Errors
1	625.61 ms	0	1047.09	4
2	383.69	2	422.34	3
3	692.19	0	694.69	2
4	607.62	0	864.93	0
5	651.68	2	689.08	1
6	800.67	1	902.67	2
7	459.48	0	467.9	1
8	808.29	0	1028.87	1
9	905.6	3	1116.02	3
10	712.09	0	861.62	3
11	731.62	0	788.9	2
12	1142.25	1	1034.37	5
13	872.34	0	1059.91	0
14	613.97	0	729.41	1
15	960.96	0	874.54	2
16	714.9	0	741.37	0

Appendix 5 – Statistical analyses

An Analysis of Variance (ANOVA) is a statistical test used for comparing means from three or more groups or conditions in an experiment. The test calculates *F*, which represents the proportion of the variance between groups or conditions to the variance within the groups or conditions. Post hoc tests consist of pairwise comparisons designed to compare all different combinations of experimental group or condition means.

A repeated measures ANOVA is a special type of the test used when the same participants take part in all conditions of an experiment. This type of design has the advantages of reducing the unsystematic variability that is typically present in a standard between-group ANOVA, and also reducing the number of participants needed.

When reporting the results of an ANOVA, the *F*-ratio is preceded by the degrees of freedom, and then followed by the significance value *p*.

Thus, for the example notation:

$F\ (3, 56) = 9.75$, $p < .001$

> The *F*-ratio is 9.75.
>
> (3, 56) are the degrees of freedom for the effect of the model (here: 3), and for the residuals of the model (here: 56), respectively.
>
> $p < .001$ is the significance value of the *F*-ratio, which here is smaller than .001, indicating the probability that an effect should occur by chance (here, smaller than 1 in 1000). Typically, effects with p-values smaller than .05 are treated as significant.

For more details on statistical tests, see Field (2013) and/or Lowie & Seton (2013).