Dáša Kadlecová
4061780

L2 collocations and the L2 mental lexicon

Master's thesis

Master's programme: Language, Mind and Society (Taal, Mens en Maatschappij)
Utrecht University
June 2014

Supervisor: prof. dr. R.W. J. Kager
Second reader: dr. Bill Philip

Word count: 13 331

Table of contents

Acknowledgements

Firstly I would like to sincerely thank my supervisor prof. dr. R.W.J. Kager for all his remarks, feedback and support and for guiding me throughout the process of writing my master's thesis. Without his help I would not have been able to show my full potential and write a master's thesis of such a standard as it is now. Secondly I am genuinely grateful to my parents and partner for all their support. Without them none of this would have been possible. Additionally I am thankful to my friend Lisette for her help and enthusiasm about my work. I am also very grateful to everyone who read it and pointed out my mistakes. Last but not least I would like to thank the third-grade students at the secondary school in Bratislava (Slovakia) for participating in the experiment.

1 Introduction

Second language (L2) collocations are an interesting phenomenon. Although knowledge of L2 collocations is significant to achieve mastery of a foreign language (Bahns & Eldaw, 1993), L2 learners usually pay little attention to (systematically) learning these structures. It has been generally acknowledged that L2 learners have problems with acquiring L2 collocations and that many factors, namely low L2 proficiency and insufficient input, may serve as constraints on the production of correct L2 collocations. Studies investigating the acquisition of L2 collocations mostly try to define to what extent L2 learners are influenced by their native language (L1) knowledge (Gabrys-Biskup, 1992; Farghal & Obiedat, 1995; Wolter & Gyllstad, 2011), whether L2 proficiency is the most influential factor in the acquisition of L2 collocations (Ha, 1988; Gitsaki, 1996; Bonk, 2000; Barfield, 2003) and what role the size and the structure of the L2 mental lexicon play in the process (Mochizuki, 2002; Gyllstad, 2007). The aim of this thesis is to examine whether the size of the L2 mental lexicon is positively correlated with knowledge of L2 collocations. Due to the fact that L2 learners with bigger L2 vocabularies are generally more proficient in an extensive range of language skills (Meara, 1996), we suggest bigger L2 vocabulary size may go hand in hand with better knowledge of L2 collocations. Besides investigating participants' knowledge of L2 collocations we aim to test whether L2 learners can apply this knowledge in order to easily acquire new L2 collocations and how is knowledge of L2 collocations influenced by interference.

The initial stage of L2 acquisition is L1 which is gradually replaced by L2 (Ellis, 1994). Accordingly, production mistakes ascribed to interference are more likely to occur in early stages of L2 acquisition. Nonetheless, even in the later stages, L1 still plays a significant role in L2 acquisition since even high proficiency L2 learners make mistakes with respect to L2 collocations so they may be influenced by their L1 knowledge. Therefore some other factors should effect the acquisition of L2 collocations. If we do not take the amount of input into consideration, we can claim the size and the structure of the L2 mental lexicon is the determining factor for the acquisition of L2 collocations. The bigger the L2 mental lexicon is, the broader network of word connections and associations it comprises, which makes it easier for L2 learners to use and acquire new L2 collocations. However adding a new item into unstructured L2 mental lexicon is more difficult with increasing L2 vocabulary size. Therefore not only the size but also the structure of the L2 mental lexicon is significant for the acquisition of L2 collocations.

This thesis is organized in four chapters. After the introductory chapter, in which the aim of the thesis is presented, the second chapter discusses collocations in more detail. It presents a detailed literature review on collocations and L2 collocations and discusses what role interference as well as the structure and the size of the L2 mental lexicon have in the acquisition of L2 collocations. At the end of this chapter the research questions together with the hypotheses are listed. The third chapter comprises the methodology of the experiment describing its participants, materials and procedure as well as the results. The final chapter focuses on discussing the results with respect to our hypotheses. The conclusion together with suggestions for future research is presented at the end of this chapter.

2 Literature review

2.1 Collocations

In order to analyse the acquisition of collocations, to be precise L2 collocations, it is significant to provide its definition in terms of linguistics. Etymologically speaking, the word collocation comes from Latin word *collocat*, which means "he places together". The Oxford English Dictionary[1] defines collocations as combinations of words that are habitually juxtaposed with a frequency greater than chance, meaning sequences of words that co-occur more often than expected. To put it differently, two words combined together create a phrase, where one word carries the meaning and the other modifies and specifies it. We may say that one member of a collocation has its regular meaning and the other, often an adjective, is idiosyncratic. To illustrate this with an example, in the collocation s*trong coffee* the noun *coffee* is regular and means only coffee and nothing else; however, the adjective *strong* adds a particular meaning to the phrase which in the case of *coffee* can only be expressed by this specific adjective. Although grammatically speaking the phrase **powerful coffee*[2] would be correct as well, however, *powerful* does not collocate with *coffee* even though *powerful* and *strong* are synonyms. Additionally the collocation *strong coffee* is a single entity even though *strong* has modified the meaning of *coffee*. In addition, the meaning of the adjective *strong* has been slightly altered as well, since the adjective *strong* primary defines muscular strength. However, this general meaning is not conveyed in the collocation *strong coffee*. Wolter (2006) described this as "lexical interaction" leading to "conceptual modification". By creating correct lexical combinations the meaning of the concept can be modified.

When various collocations are compared, we may notice that based on the analogy between such expressions, combining words together to create correct collocations does not necessarily have to be unpredictable. Although there are no rules about which words collocate together, a certain pattern may be observed. As an example, the adjective *strong* expressing the "strength" of a drink can be found in various collocations such as *strong coffee*, *strong tea*, *strong beer*, *strong alcohol*, etc. We may claim that if an adjective is combined with a noun, then the adjective can also be combined with other nouns belonging to the same group. In the case of the above-mentioned example, the adjective *strong* can be combined with other nouns belonging to the group "drinks". Or the adjective *heavy* collocates with the noun *rain*, thus we may expect it will also collocate with other nouns expressing "weather" such as *heavy*

[1] Online version, retrieved from http://www.oxforddictionaries.com/definition/english/collocation?q=collocation
[2] In this thesis the symbol "*" is used to identify incorrect collocations.

snow, heavy blizzard, etc. This similar pattern, as we will refer to it, can be observed not only in adjective-noun collocations but also in other collocations types e.g. verb-noun collocations. For instance the verb *lose* collocates with nouns expressing a person's character such as *lose patience, lose temper, etc.* We have to mention that applying the same analogy when creating word combinations does not always result in correct collocations. Even though the noun *wind* belongs to the previously mentioned group "weather" the collocation **heavy wind* is incorrect in English. Nonetheless we may hypothesize the acquisition of new collocations may be facilitated by knowledge of the presence of such patterns.

As one of the few researchers who focused on collocations, Cowan (1989) tried to distinguish collocations from idioms, providing characteristics of collocations, saying they consist of more than one word and they are resistant to lexical substitution. Firstly these features apply to idioms as well; secondly she also mentioned that in certain cases a part of a collocation can be replaced by a synonym, for example *nasty/ugly/messy divorce* (it is possible to combine the noun *divorce* with three adjectives to express the same concept). She concluded that there is yet no perfectly reliable definition to distinguish collocations from idioms; however, some features may be used as hints to discriminate between them. Howarth (1998) also tried to draw a line between collocations and idioms. Among other things in his work he criticized other researchers for using the terms collocations and idioms interchangeably and not paying attention to the fact that under a closer examination they significantly differ from each other. He claimed that in current models the division of word combinations into idiomatic and non-idiomatic was not sufficient enough and further sub-categorization was necessary. Therefore in order to make a clearer division, Howarth proposed a new categorization of these expressions referred to as collocational continuum, namely into free combinations (for instance *under the table*), restricted collocations (*under attack*), figurative idioms ([*to be] under the microscope*) and pure idioms (*under the weather*). He explained that free combinations are often referred to as "open" or "free" collocations, comprise units used in their literal senses (i.e. have fully compositional meanings) and are freely commutable. Restricted collocations have one element, usually an adjective, a preposition or a noun, which often has a figurative meaning that can only be understood in the context of a limited number of collocates. Moreover Howarth illustrated the difference between figurative and pure idioms. The former have a metaphorical meaning as well as a current literal interpretation; the latter have a unitary meaning that cannot be derived from meanings of its individual parts. This collocational continuum was created by applying criteria such as semantic specialization,

idiomaticity and restricted collocability, each of which is gradable. Semantic specialization indicates that the semantic meaning of a collocation is specific and when parts of a collocation are changed, its meaning is modified as well. Idiomaticity can be explained as unpredictability of the meaning of a word combination from the meaning of its individual constituents (Glaser, 1998). Restricted collocability comprises limited combining of words together and blocking of lexical substitution. Additionally, Howarth (1998) claimed that some restricted collocations are on the borderline with figurative idioms, making the system of categorization of these items unreliable.

In order to discriminate between collocations and idioms in this thesis we will follow Howarth's model; nonetheless we believe that the most significant criterion to differentiate between collocations and idioms is the predictability or compositionality of the meaning of an expression. We suggest that the meaning of a collocation is the sum of the meanings of its parts, therefore it can be partially predicted. To illustrate this with an example, the idiom *paper tiger* has nothing to do with neither paper nor tiger, its meaning must be understood metaphorically (something or someone who appears to be threatening but in reality is not). On the other hand the collocation *hot summer* contains a lexical item whose literal meaning contributes to the overall meaning of the expression next to an element of which the specific meaning in this context is not fully predictable, for instance the adjective *hot*. Although based on knowledge of the general meaning of the adjective *hot* the meaning of *hot summer* can be predicted, however the range of hotness differs in the collocation *hot coffee* and *hot summer.* To sum up, the term collocation(s) will be used in this thesis to define a combination of typically two words (e.g. a verb and a noun or an adjective and a noun), out of which the modifying part (e.g. an adjective) usually cannot be replaced by a synonym and their meaning is partially predictable. Furthermore, as mentioned above the semantic meaning of the combined words does not help us in defining whether such combination is acceptable or not since the combining is arbitrary. To put it differently, there is no simple rule that can tell us which words collocate together. This brings us to the generally acknowledged notion that it is exceedingly difficult for L2 learners to acquire L2 collocations due to unpredictability of combining specific words together. However, we may claim that previously acquired vocabulary may help in acquiring new L2 collocations. This issue will be discussed more deeply in the following sections.

When investigating the phenomenon of collocations only few studies concentrate on collocations without discussing L2 or L2 collocations. As mentioned before such studies

mostly try to define the difference between idioms and collocations (Cowan, 1989; Howarth, 1998) or simply define collocations in general (Siepmann, 2005; Shin, 2006). In order to answer the questions concerning the elements that form collocations and whether they are arbitrary, Siepmann (2005) suggested a new division of the spectrum of collocations and tried to broaden the definition of collocations. He concluded that the future studies concerning lexicography must necessarily concentrate on the phenomenon of collocations since they were neglected in the past. Furthermore, Shin (2006) concentrated on investigating the usage of collocations. Among other things he found out that collocations are more frequently used in spoken language than in written language and that the length of collocations has an impact on the frequency of usage, claiming shorter collocations are used more often than longer ones. Lin (1999) tried to determine whether a collocation is compositional or not. In his work he presented a method that enables to identify non-compositional phrases given the assumption that non-compositional expressions have clearly different mutual information value than those that are similar to their literal meanings. As one of the few Jervic (2011) used the collocational behaviour as a factor for determining whether synonyms can truly be absolute. After comparing the collocational behaviour of 36 pairs of synonyms she concluded the concept of absolute synonyms is unrealistic.

2.2 L2 collocations

When studying second language[3] acquisition (SLA), there are many factors that play a role and influence the process of SLA, namely the amount of input, motivation, age of onset, native language (L1), etc. (Moyer, 1999; Dörnyei, 2003; Unsworth, 2008). Most of the researchers focusing on L2 collocations try to define the importance of factors such as L1 (Gabrys-Biskup, 1992; Farghal & Obiedat, 1995; Wolter & Gyllstad, 2011), L2 proficiency (Ha, 1988; Gitsaki, 1996; Bonk, 2000; Barfield, 2003) or L2 vocabulary size (Mochizuki, 2002; Gyllstad, 2007) and see if there is any correlation between knowledge of L2 collocations and above-mentioned factors. Regardless of the tested variables, all studies concerning L2 collocations have come to the conclusion that acquiring L2 collocations is extremely difficult for L2 learners. Even though many factors have effect on the process, Wray (2002) claimed that one of the main causes why L2 collocations are difficult to acquire may be the way in which L2 learners acquire new words. She suggested that when compared

[3] In this thesis the terms "second language" and "foreign language" are used interchangeably.

to natives, who see collocations as units of words, L2 learners acquire individual words of a collocation separately. We may say that such a statement is not completely accurate, as natives also acquire their native language word by word, which means each of the words of a collocation must be stored in their mental lexicons separately. Yet the major difference between L1 and L2 learners is the fact that L2 learners fail to store word combinations as single units. Although they are familiar with all words in a collocation, they might still fail to produce the correct L2 collocation (Wolter, 2006). To illustrate this with an example, the collocation *heavy rain* comprises an adjective and a noun. When learning English, either L1 or L2 learners acquire words *heavy* and *rain* individually and store them in their mental lexicons separately. Later on L1 learners learn the collocation *heavy rain* and store it in their mental lexicons also as one single unit. Even though L2 learners might be familiar with this collocation and the whole word combination may be stored in their L2 mental lexicons as a single unit, the associations between the concept and individual words of the collocation are stronger and thus preferred over the association between the concept and the whole collocation as a single unit. As mentioned by Wray (2002), L2 learners have problems with establishing strong associations between words that form collocations. Therefore we may propose that the structure of the L2 mental lexicon definitely plays a significant role in learning L2 collocations.

Although knowledge of L2 collocations is significant to achieve mastery of a foreign language (Bahns & Eldaw, 1993; Chan & Liou, 2005), students are often not aware of target language collocations as a potential problem and they are not used to paying attention to such structures. Schmitt (1999) highlighted that collocational behaviour of a word is one of eight competencies (i.e. spoken and written form of the word, its conceptual meaning, collocational and grammatical behaviour, its frequency, associations and stylistic constraints) included in what is known as word knowledge, which defines what it means to have a comprehensive knowledge of a word on a native-like level. Since L2 learners make production errors with respect to L2 collocations, we may suggest that L2 learners do not acquire words with all of their competencies. Moreover, due to the unpredictability of which words collocate together, L2 learners may not easily acquire collocations (Bahns & Eldaw, 1993; Chan & Liou, 2005). Nevertheless this might not be the only explanation. As generally acknowledged, the amount of input plays a significant role in L2 acquisition. Due to the fact that L2 learners find individual words of collocations in the input more often than whole word combinations, the associations between separate units of collocations and concepts are stronger and therefore

preferred over associations between whole collocations and concepts. Similarly Chan & Liou (2005) claimed that frequency may affect the acquisition of L2 collocations. To put it differently, L2 learners take advantage of the tremendous amount of input to strengthen their associations between concepts and L2.

It has been generally acknowledged that when L2 learners acquire L2 words there is no direct link between the concept and the appropriate L2 word (Kroll, 2002), meaning L2 learners are at first unable to use L2 words without mediation through L1. The revised hierarchical model (see Figure 1) designed by Kroll & Stewart (1994) merges the word association and concept association model. The word association model proposes that L2 words are directly connected to their L1 equivalents, while according to the concept association model L2 words are connected to their meanings without L1 playing any significant role. Concepts can be expressed by both L1 and L2 word(s); however, conceptual links between concepts and L1 word(s) are much stronger (Kroll, 2002). Therefore when confronted with the need or the requirement to use L2 word(s), L2 learners use L1 as a mediator and try to find L2 equivalents for already known L1 word(s). In Figure 1 the solid and dotted lines represent different strengths of relationships between concepts and L1 or L2 words. While the lines between concepts and L1 or L2 are bidirectional, meaning the learner uses conceptual links in both directions, lexical links work only in one direction. As mentioned in Kroll (2002) the connection from L1 to L2 is not particularly strong as learners do not use L2 in this way. On the contrary, the connection from L2 to L1 is crucial especially at the beginning of SLA since learners have no other option to express concepts given the fact that conceptual links between L2 and concepts are not solid enough. Kroll (2002) also claimed L2 proficiency has an effect on whether conceptual links between L2 and concepts are strong enough and thus preferred over the other links with mediation through L1. We may propose that when acquiring and using L2 collocations L2 learners also use L1 as a mediator; nonetheless whether links between L2 and concepts are strong enough depends not only on L2 proficiency but also on L2 vocabulary size.

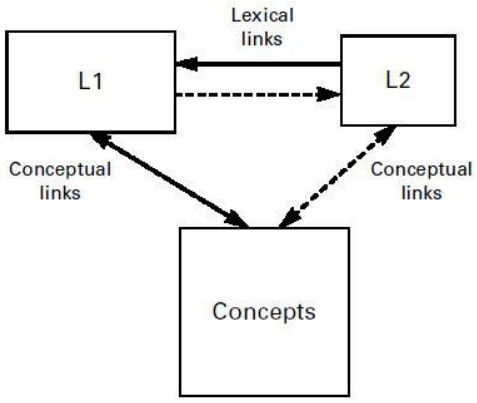

***Figure 1**: Revised hierarchical model*

Based on the revised hierarchical model we may predict that similarities or differences between two languages may determine how fast L2 learners start to use conceptual links between L2 and concepts regarding L2 collocations. If two languages are similar, L2 learners might not even notice how quickly they stop using L1 as a mediator as the strength of conceptual links between L2 and concepts are rather great right away. However, this transition may last longer if the languages are too different from each other and it is thus difficult for L2 learners not to use L1 as a mediator. Furthermore, when collocations are not used to express concepts in L1, L2 learners might be forced to develop conceptual links between L2 and concepts more quickly. A similar scenario may be expected when collocations only occur in L1 and not in L2. Additionally, we may predict that learners who still prefer using L1 as a mediator may be more influenced by negative transfer and thus make more mistakes attributed to interference.

Referring back to studies on L2 collocations, Chan & Liou (2005) investigated the influence of web-based learning of English noun-verb collocations of Chinese college students in Taiwan. Their aim was to determine whether the computer programme used in the study can help L2 learners to improve their knowledge of L2 collocations and if the effect is permanent. During the experiment five tests (pre-tests and post-tests) were administered. The results indicate that students made significant improvement in their knowledge of L2 collocations; however, the improvement was only temporary since the scores on the post-test were lower than on the last pre-test but still above their entry scores. Chan & Liou proposed that the decline can be attributed to the lack of input, as the students did not have much input concerning L2 collocations and thus simply forgot them. They concluded that students with the lowest scores on the first test made the most significant progress. Bonk (2000)

investigated students of English as second language (ESL) of a wide range of proficiency levels. He concentrated on determining the correlation between knowledge of L2 collocations and general L2 proficiency. In addition to the major purpose of the study, he tried to find out whether there is a correlation between knowledge of L2 collocations and the length of residence in an English speaking country. Together with a simplified version of the TOEFL test to measure the general L2 proficiency, a collocation task consisting of fill-in as well as multiple choice questions was administered. He found that there is indeed a correlation between knowledge of L2 collocations and L2 proficiency, even though in other studies no relationship has been established at all (Barfield, 2003). However, no significant correlation was found between knowledge of L2 collocations and the length of residency in an English speaking country. Moreover he concluded that although individual differences that occurred in the study cannot always be predicted from the differences in the proficiency level, they may be attributed to factors such as memory. Similarly Ellis (1996) claimed short-term memory capacity may serve as a constraint on acquiring new L2 collocations and L2 in general.

The most common strategy L2 learners use when producing L2 collocations was described by Huang (2001), namely that L2 learners rely heavily on L1 knowledge as the only resource. In addition, many studies focusing on L2 collocations have illustrated that L2 learners score better on collocations that have L1 equivalents (Gabrys-Biskup, 1992; Huang, 2001). This observation may seem obvious; nonetheless it shows that the revised hierarchical model depicted in Figure 1 also applies to collocations since L2 learners use L1 as a mediator when producing L2 collocations. We may hypothesize that most of the L2 learners start acquiring a language with the assumption that there is one-to-one correspondence between their L1 and L2 and do not make use of conceptual links between concepts and L2; however, such supposition may last longer in the case of collocations, especially those used rather infrequently as the amount of input often plays a major role in SLA and thus may determine the strength of the association between concepts and L2 collocations (Chan & Liou, 2005).

Acquiring new L2 words is comparatively easier than acquiring new L2 collocations. Paradigmatic and syntagmatic connections can account for the difference in the difficulties. Words that are paradigmatically related can substitute each other in a sentence without considerably modifying its meaning (e.g. *horse* → *animal*); syntagmatic connections represent a co-occurring relationship between two words creating a phrase (e.g. *man* → *tall, walk*). Since the nature of categorization of *horse* belonging to the group *animal* remains, L2

learners just need to substitute L1 words for L2 words in their L2 mental lexicons. Therefore paradigmatic connections between L2 words are more easily established as the nature of the connections between words and concepts are not changed. On the contrary syntagmatic connections are involved in conceptual modification (Wolter, 2006). Due to the fact that syntagmatic connections creating word associations vary from language to language; L2 learners have to make new syntagmatic connections between individual words (*influence* → *wide*, ~~*heavy*~~) and gradually build new networks in their L2 mental lexicons.

2.3 Interference

In linguistics interference[4] is described as applying L1 knowledge to L2 (Richards & Schmidt, 2002). Two types of language transfer may occur in the acquisition process, namely positive (if L1 and L2 can be defined as similar languages, SLA can be facilitated by L1 knowledge) and negative (dissimilarity of the languages may cause impeding of SLA). Although most of the studies concerning interference focus on transfer from the native language to the target language, it is generally acknowledged that transfer can work from L2 to L1 as well as from a second language to another foreign language(s) (Richards & Schmidt, 2002). Furthermore, Ellis (1994) explained in his book that the starting point of SLA of every L2 learner is L1, which is gradually replaced by the target language throughout the process of SLA. Accordingly the errors attributed to transfer are more likely to occur in early stages rather than in later stages of the acquisition process. In addition, Krashen (1981) pointed out that learners may substitute an L2 segment for an L1 based utterance if they have not yet acquired enough L2 knowledge due to the lack of input. He also claimed that interference seems to be stronger and more observable if the L2 is acquired without any access to natural setting and L2 learners have received only instructional input without being in contact with native speakers.

When discussing interference, many linguists used to refer to Lado's formulation of Contrastive Analysis Hypothesis appearing in his book Linguistics Across Cultures (1957), which basically predicts a strong correlation between the differences of L1 and L2 and the difficulty to acquire L2. In other words, elements of languages that are similar barely pose any difficulties for L2 learners, unlike dissimilar elements which make SLA more difficult. The hypothesis was analysed and practiced in the 1950s and 1960s (Richards & Schmidt, 2002). Many critics have pointed out that the difficulties predicted to occur by Contrastive Analysis

[4] In this thesis the terms "interference" and "transfer" are used interchangeably.

(CA) do not occur in the actual performance of L2 learners, likewise many errors made by L2 learners are not predicted by CA (Sridhar, 1975). The CA was therefore replaced by so the called Error Analysis (EA) as the alternative in the 1970s, which demonstrates that not all mistakes can be attributed to the interference of L1 (Richards & Schmidt, 2002). In order to put the analysis into practice a classification of errors was needed. Thus a general distinction was made between interlingual (results of partial or incorrect learning of L2) and intralingual errors (divided into communication-based errors, errors of avoidance, overgeneralizations, errors of overproduction, developmental errors, induced errors and simplifications) (Richards & Schmidt, 2002). Since the cause of the incorrect usage of L2 elements cannot always be categorized, EA seemed to be insufficient. Later on Interlanguage Analysis was introduced in the late 1970s, which claims that the grammatical system used by L2 learners during the acquisition process differs from both L1 as well as L2 (Huang, 2011; Richards & Schmidt, 2002).

As we have discussed above, if L1 and L2 are perceived by L2 learners as similar languages, SLA may be facilitated. This positive transfer is sometimes compared to a U-shaped acquisition process (Ellis, 1994), where at first L2 learners acquire the target language very easily, later after making mistakes they realize that sometimes there is no one-to-one correspondence, and at last they return to the correct usage of L2. Even though only a few studies concentrate on effects the positive transfer has on SLA (Huang, 2011), many researchers try to focus more deeply on the question to what extent production errors of L2 learners are the result of interference or simply general language acquisition process. In his book Ellis (1994) mentioned various results of different empirical studies concerning the factor of interference in SLA. He highlighted the enormous variety in the results that ascribe errors to L1 transfer. Interestingly, the range goes from 3 percent (in Dulay & Burt, 1973) to 51 percent (in Tran-Chi-Chau, 1975). Ellis (1994) concluded that discriminating the two factors is crucial and he agreed with the general opinion that interference works in complex ways. To put it differently, it represents only one of many processes involved in SLA. Ellis also suggested that L2 learners know that certain L1 segments are possibly transferable while others not; however, each L2 learner understands this distinction differently and decides which segments he/she will actually transfer to L2. Additionally, during the acquisition process L2 learners simply make wrong assumptions, thinking there is a one-to-one correspondence between L1 and L2 (Farghal & Obiedat, 1995; Huang, 2001). In other words, L2 learners expect the properties of their mother tongue to appear in L2 as well.

Returning to the topic of L2 collocations, several studies investigating L2 collocations have attributed the mistakes made by participants to interference (Farghal & Obiedat, 1995; Gabrys-Biskup, 1992; Gyllstad, 2007; Wolter & Gyllstad, 2011). For instance Gyllstad (2007) described the difference in the results of high and low proficiency learners, stating low proficiency learners are more likely to rely on their L1 knowledge and make mistakes explained by negative transfer. In addition, he stated that one of the causes may be the difference in the amount of input between low and high proficiency learners, claiming high proficiency learners have received enough input to realize that a one-to-one correspondence is not always the case and therefore make less or even no mistakes that can be attributed to L1 transfer. Among others Gabrys-Biskup (1992) examined Polish students of English and their knowledge of L2 collocations. Even though he stated that when asked to provide Polish equivalents to English collocations the participants' answers were hundred percent correct, the author did not present any figures on the correct answers of the second part of the test, in which participants were asked to translate Polish collocations into English. Gabrys-Biskup claimed the majority of L2 errors are to be attributed to L1 interference. He also concluded that L2 learners tend to rely on L1 knowledge when they do not know or they are unsure of the correct answer. Krashen (1981) suggested that such a strategy used by L2 learners can be explained by insufficient input. We may hypothesize that the fact that many L2 learners rely on L1 knowledge of collocations can be explained by lack of learners' awareness of the existence of L2 collocations.

To sum up, we expect negative transfer to be displayed in two manners – a) due to the lack of input or b) due to the incorrect assumption about a consistent one-to-one correspondence between L1 and L2, L1 is literally translated to L2. In addition, we may propose that L2 learners are more easily influenced by their knowledge of L1 collocations if a tempting example is presented as one of the possible collocation structures. As mentioned above, the starting point of every L2 learner is the L1, meaning the influence of L1 should gradually decrease during the acquisition process. Moreover, due to various factors such as quality and quantity of input, L2 proficiency and vocabulary size L2 learners become aware of the non-transferability of L1 features to L2. As highlighted by Ellis (1994) it is important to investigate how L1 knowledge interacts with other factors.

2.4 L2 vocabulary size and the structure of the L2 mental lexicon

The mental lexicon can be defined as a mental representation of a dictionary L2 learners use during SLA. Such lexicon comprises all information with respect to acquired words such as spoken and written form of the word, its conceptual meaning, grammatical and collocational behaviour, its frequency, associations and stylistic constraints, altogether creating word knowledge as proposed by Schimitt (1999). However, the explanation that can account for L2 learners having problems with for instance L2 collocations may be the fact that they do not organize and structure their L2 mental lexicons according to word knowledge competences.

Even though the idea that the structure of the L1 and L2 mental lexicon differ from each other has been generally acknowledged, Wolter (2001) tried to prove that the L2 mental lexicon of a non-native speaker is in fact quite similarly structured as the mental lexicon of a native speaker. Although his study actually confirmed the already accepted notion about the dissimilarity of the two lexicons, he claimed the L2 mental lexicons were not as loosely structured as in comparison to L1 mental lexicons as other researchers had suggested. Among others Zareva (2007) compared the mental lexicons of L2 learners and native speakers. Participants were investigated by means of a word association test, where they were asked to respond to each stimulus with the first word that came to their mind. She found a significant difference in quantitative and qualitative characteristics of lexical organization. She concluded that meaning connections of advanced learners resembled native speakers' quantitative patterns of word associations (the author defined quantitative measures as strength of response, number of responses, response commonality and heterogeneity, etc.); whereas no similarity was established between the responses of intermediate learners and native speakers.

Even though several studies on the L2 mental lexicon focus on developing new versions of vocabulary level and size tests (Laufer & Nation, 1999; Schmitt, Schmitt & Clapham, 2001), only few studies concerning L2 vocabulary size actually discuss L2 collocations and a possible positive correlation between L2 vocabulary size and knowledge of L2 collocations (Bahns & Eldaw, 1993; Gyllstad, 2007). On one hand Bahns & Eldaw (1993) concluded that knowledge of L2 collocations does not go hand in hand with knowledge of general vocabulary, saying there is no significant correlation between the two features as knowledge of general vocabulary of ESL students was more profound than their knowledge of English collocations in the administered test. On the other hand results of Gyllstad (2007) show exactly the opposite; namely a positive correlation was observed since scores on the test highly correlated with the scores on the receptive vocabulary size test administered to

Swedish learners of English. Furthermore, Wolter & Gyllstad (2011) tested collocational links in the mental lexicon of English learners of Swedish in various conditions, namely collocations with and without L1 equivalents. They investigated whether L2 learners acquired L2 collocations that have an equivalent in their native language differently than collocations that only occur in L2. To test their hypotheses a primed lexical decision task was administered. Firstly only a few pieces of information with respect to the proficiency levels of participants were provided. Since the authors defined only the average of self-evaluation proficiency scores (e.g. 6.8 for speaking, 6.1 for writing), we do not know if any beginner or intermediate learners took part in the study. Secondly we may point out that no list of stimuli administered in the study was provided in the article. Nonetheless the results illustrate that L2 learners can recognize as well as process L2 collocations that also occur in their native language more effectively than L2 collocations without L1 equivalent. Thus they concluded that L1 may have a considerable influence on the development of knowledge of L2 collocations. Moreover L2 collocations without L1 equivalent should according to Wolter & Gyllstad be immediately stored in the L2 mental lexicon as L2 collocations without L1 playing any role. However, we may claim that such an assumption is not necessarily true as L2 collocations are also stored individually as two separate words in L2 learners' mental lexicons.

Although agreeing with the assumptions made by Gyllstad (2007) and recognizing the importance of input and a possible correlation between L2 proficiency and knowledge of L2 collocations, we propose that vocabulary size, namely the size of the L2 mental lexicon, plays a significant role in acquiring L2 collocations. We believe that the bigger L2 mental lexicon is, the broader network of word connections and associations it comprises. Therefore L2 learners with bigger L2 vocabularies should more easily acquire L2 collocations due to their structured networks in the L2 mental lexicon. Given the fact that L2 mental lexicons are generally smaller than L1 mental lexicons, Meara (1996) proposed that an item in the L2 mental lexicon may be directly connected only to a small number of words creating word associations, unlike items in the L1 mental lexicon that have a significantly higher number of associations. Moreover he claimed that at a certain point during the acquisition process the size dimension of L2 vocabulary becomes less important as the significance of the structure and organization of the L2 mental lexicon increases. Adding a new item into an unstructured lexicon clearly becomes more difficult with increasing L2 vocabulary size. We may suggest that one of the causes of L2 learners having problems with acquiring new collocations is the

fact that they usually do not acquire words systematically with all the competences of word knowledge as proposed by Schmitt (1999). Therefore the made associations are not as strong as in comparison with a native speaker (Wray, 2002). Thus we may conclude that the structure of the L2 mental lexicon is a crucial factor in acquiring new L2 collocations. We may even assume that the already acquired collocations may help acquiring new collocations, since L2 learners may be aware of certain patterns occurring in word combinations. To illustrate our assumption with an example, knowledge of collocation *lose patience* may make the process of learning similar collocations as *lose temper* and *lose faith* easier.

2.5 Research questions and hypotheses

The purpose of this thesis is to find out to what extent is the size and structure of the mental lexicon of L2 learners correlated to their knowledge of L2 collocations. Bearing previous studies in mind we propose an experiment investigating Slovak intermediate learners of English and their knowledge of English collocations. The designed experiment will test whether participants with varying vocabulary sizes perform differently on the test of L2 collocations. More specifically we will investigate: a) correlation of general L2 vocabulary size and knowledge of L2 collocations, b) to what extent does the size of the L2 mental lexicon help acquiring new collocations with similar patterns (as described in the first section) and c) the influence of interference on knowledge of L2 collocations. We will try to find out whether a learner who has already acquired frequently used collocation *pay attention* will be able to use this knowledge when acquiring new rather infrequent collocation *pay respect*. Our study questions are defined as follows:

> Q1: Do Slovak learners of English with higher scores on the vocabulary test also achieve better scores on the collocation task?
>
> Q2: Do already acquired L2 collocations help Slovak learners of English to acquire new collocations with similar patterns?
>
> Q3: Do Slovak learners of English with higher scores on the vocabulary test make fewer mistakes attributed to interference than learners with lower scores?
>
> Q4: Are Slovak learners of English more influenced by their knowledge of frequently used Slovak collocations rather than infrequent Slovak collocations?

Based on the knowledge of results from the previous studies and of the available literature with respect to this phenomenon we expect to find a positive correlation between scores on the vocabulary task and scores on the collocation task and fewer errors ascribed to interference. In other words, due to the fact that the major dimension of lexical competence is size and L2 learners with bigger vocabularies are usually more proficient in an extensive range of language skills (Meara, 1996), we expect better score on the vocabulary task will go hand in hand with better score on the collocation task and fewer mistakes attributed to transfer. As discussed in the previous sections of this thesis mistakes attributed to interference are more likely to occur in earlier stages of SLA, therefore we suggest that the amount of acquired vocabulary, in other words vocabulary size, may affect the transfer error rate, saying that learners with bigger L2 mental lexicons will be aware of the non-transferability of L1 into L2 and therefore make fewer errors in the collocation task attributed to interference. In addition, we believe frequency might be a significant factor with respect to interference, suggesting that participants will be more influenced by their knowledge of frequently used Slovak collocations. Furthermore, we expect already acquired L2 collocations may help an L2 learner to learn new L2 collocations with similar patterns. Our hypotheses are stated as follows:

> H1: Participants with higher scores on the vocabulary test will have better scores on collocation task than participants with lower scores on the vocabulary test.
>
> H2: Participants with knowledge of rather frequently used L2 collocations will also be able to identify rather infrequent collocations with similar patterns as correct.
>
> H3: If mistakes in the collocation task ascribed to interference occur, participants with higher scores on the vocabulary test will make fewer such mistakes than participants with lower scores on the vocabulary test.
>
> H4: If mistakes in the collocation task ascribed to interference occur, participants will be more influenced by their knowledge of rather frequent Slovak collocations than infrequent ones.

3 Experiment

3.1 Participants

For the purposes of this experiment we tested a group of Slovak secondary school students (*N*= 30, *n* males = 21, *n* females = 9) attending the same secondary school in Slovakia. Even though the age ranged from 17 to 19 years old (M_{age} = 17.83, *SD* = 0.53), all participants were in their third year of studying at the secondary school. Due to the fact that the syllabus as well as the study material for English classes is identical, the gained input after three years of studying at the secondary school should be identical as well. Furthermore, participants all started learning English at an elementary school with an exception of 7 participants who began learning English at the age of 5 (before starting attending an elementary school). However, none of them were raised as a bilingual; in other words, all participants considered English as their second language, thus we do not consider the difference in the age of onset significant. Moreover, although the participants' length of exposure to English varied ($M_{length\ of\ exposure}$ = 9.6 years, *SD* = 2.86), we may hypothesize that based on knowledge of the Slovak educational system, the difference in the length of exposure does not matter significantly as most of the students coming to a secondary school have only basic knowledge of English and only after 4 years of studying at a secondary school they have to pass a school leaving exam in English on B1 or B2 level. It would not have been possible to find a more homogeneous group of secondary school students who attended the same elementary school. Last but not least the proficiency in English as reported by the participants ranged from B1 to B2 level (*n* B1 = 11, *n* B2 = 19); however, we may generalize them as intermediate learners of English. For a more detailed table of participants see Appendix 1.

3.2 Materials

For the purposes of the experiment we decided to use only verb-noun and adjective-noun collocations due to rather frequent usage of these word classes in course books and everyday speech. Even though the course book used by the participants during English lessons does not mention the existence of collocations, it comprises few exercises concerning word combinations. However, the chosen collocations were not explicitly mentioned in any of such exercise. Overall we have tested 15 verb-noun and 12 adjective-noun frequent as well as rather infrequent English collocations (27 in total) as depicted in Table 1. The frequency was computed by using Google search, the date of collection being 20th April 2014. When tested

for frequency each collocation was written with quotation marks. In addition, to test our second hypothesis the tested collocations had to comprise structures with similar patterns that can enable L2 learners to acquire new and less frequent L2 collocations more easily. Out of 27 tested collocations, 8 verb-noun and 8 adjective-noun collocations were selected for this purpose (see Table 2). Other collocations also meet the criteria; however due to the fact that literal translation of these collocations into Slovak is correct in Slovak as well, we would not have been able to define whether the participants identified infrequent collocations using an analogy or simply relying on their L1 knowledge. The correctness of each English collocation was consulted with *Oxford Collocations Dictionary for Students of English*. Moreover 15 possible Slovak collocations (8 verb-noun and 7 adjective-noun collocations) as depicted in Table 3 were "hidden" in the collocation task. To put it differently, the participants were confronted with possible Slovak collocations translated to English that are grammatical in Slovak but incorrect in English (later mentioned as "Slovak collocation(s)"). The correctness of Slovak collocations has been consulted with an online version of the Slovak dictionary available at http://www.slex.sk and their frequency of usage in Slovak was calculated by using Google search as well, the date of collection being 20th April 2014. Each collocation was written with quotation marks when tested for frequency.

***Table 1**: Tested L2 collocations ordered by frequency of occurrence on the Internet.*

VERB-NOUN Collocations	frequency
Make an effort	162 000 000
Pay attention	17 200 000
Give thanks	3 010 000
Pay respect	1 500 000
Lose hope	1 230 000
Give a compliment	1 180 000
Lose patience	649 000
Pay a compliment	483 000
Give attention	435 000
Lose balance	217 000
Hold hope	109 000
Lack patience	77 100
Lack courage	67 900
Lose temper	56 000
Give congratulations	13 000

ADJECTIVE-NOUN collocations	frequency
Heavy snow	6 550 000
Heavy rain	1 750 000
Wide recognition	664 000
Dense forest	458 000
Strong tea	409 000
Strong coffee	394 000
Hard punch	278 000
Thick pad	243 000
Dense wood	173 000
Wide influence	138 000
Thick eyebrow	21 400
Heavy eyebrow	4 200

***Table 2**: Tested L2 collocations with similar patterns listed with frequency of occurrence on the Internet.*

ADJECTIVE-NOUN Collocations	frequency
Dense forest	458 000
Dense wood	173 000
Heavy snow	6 555 000
Heavy rain	1 750 000
Strong tea	409 000
Strong coffee	394 000
Wide recognition	664 000
Wide influence	138 000

VERB-NOUN Collocations	frequency
Give thanks	3 010 000
Give a compliment	1 180 000
Give congratulations	13 000
Lose patience	649 000
Lose temper	53 000
Pay attention	17 200 000
Pay respect	1 500 000
Pay a compliment	483 000

***Table 3**: Slovak collocations with their Slovak forms in brackets ordered by frequency of occurrence on the Internet.*

VERB-NOUN collocations	frequency
* *Give an effort* *(dať si námahu)*	39 500
* *Hold balance* *(udržať rovnováhu)*	13 900
* *Lose courage* *(stratiť odvahu)*	5 930
* *Lack temper* *(postrádať trpezlivosť)*	5 780
* *Lose speech* *(stratiť reč)*	2 190
* *Lack balance* *(postrádať rovnováhu)*	1 650
* *Make a compliment* *((u)robiť kompliment(y))*	839
* *Lack hope* *(postrádať nádej)*	705

ADJECTIVE-NOUN collocations	frequency
* *Heavy influence* *(silný vplyv)*	121 000
* *Hard wood* *(tvrdé drevo)*	41 300
* *Thick snow* *(hustý sneh)*	31 200
* *Strong punch* *(silný úder)*	11 200
* *Wide pad* *(široká podložka)*	7 280
* *Thick rain* *(hustý dážď)*	6 690
* *Hard operation* *(ťažká operácia)*	3 390

3.3 Procedure

The experiment was carried out at the participants' school during their lesson of physics. The teacher was present in the class, however, the communication with the participants was

maintained only with the researcher who gave instructions and was present during testing for eventual questions. The instructions for all parts of the experiment were provided in English. We expected the participants to have sufficient knowledge of English to understand instructions correctly. Additionally, the participants were not informed beforehand about the experiment taking place, they were not familiar with its purposes and did not receive any kind of reward in a form of a good mark for participating. The experiment consisted of three parts: collocation task, vocabulary task and a sociolinguistic questionnaire. All three tasks were administered on several sheets of paper provided at once, on which the participants wrote their answers. Due to this fact the participants were able to rewrite their previous answers; however, each task had a time limit (10 minutes per task) in order for the participants not to overthink their answers. Overall two versions of the test with different orders of tables in the collocation task and groups of words in the vocabulary task were provided. The reason was to prevent the participants from cheating, as the experiment was administered in a class and the participants were seated next to each other.

Firstly the collocation task was presented. It was adopted from Gyllstad (2007); however, the task was adjusted to the purposes of this experiment and the used collocations were not copied from Gyllstad's experiment. The task comprised four tables, wherein verbs or adjectives had to be matched with provided nouns creating possible collocations (see Appendix 2). In each table there were three verbs or adjectives on the left listed in alphabetical order and six nouns on the top in alphabetical order as well. If a verb-noun or adjective-noun combination was according to a participant correct, he/she was asked to put a cross (X) in the cell where the verb and the noun or the adjective and the noun came together. Out of all possible combinations, few corresponded to correct English collocations and few to literal translations of Slovak collocations that were used to test the interference of L1. Furthermore, in order for the participants to get familiar with the task an example of a table with correct answers that later did not occur in the task was presented at the beginning of the test (also depicted in Appendix 2). The participants had 10 minutes to finish this task.

Secondly a vocabulary task was administered. It was a version of Vocabulary Levels Test adopted from Nation (2001), since such tests are used to give an estimate of general vocabulary size of L2 learners. The participants were tested on their knowledge of 20 groups of words (10 from the 5000 word level and 10 from academic vocabulary). The participants were asked to match meanings on the right with the correspondent words on the left as depicted in Figure 2. Words on the left were ordered alphabetically and the meanings on the

right in order of increasing length. Each meaning corresponded to only one word. Similarly to the previous task an example was presented at the beginning of the task. The time limit for this task was 10 minutes.

1. *business*
2. *clock* ___6__ *part of the house*
3. *horse* ___3__ *animal with four legs*
4. *pencil* ___4__ *something used for writing*
5. *shoe*
6. *wall*

Figure 2*: The example presented at the beginning of the vocabulary task.*

Last but not least a sociolinguistic questionnaire was administered to get information about the participants' language background, asking about the age of onset, the length of exposure, quantity and quality of input, attitude to using English, knowledge of other foreign languages, etc (see Appendix 3).

3.4 Results

The answers provided by the participants were scored by hand by the researcher. Illegible answers as well as no answers provided were evaluated as mistakes. Answers in the collocation task were evaluated as correct, incorrect or incorrect and ascribed to L1 transfer; responses from the vocabulary task were evaluated as either correct or incorrect. Each correct answer represented 1 point. The maximum score on the collocation task was 27 points (15 points for verb-noun collocations and 12 points for adjective-noun collocations); the highest score possible on the vocabulary task was 60 points. Out of 42 handed tests, only 30 were accepted for further analysis as 7 participants failed to finish both tasks and 5 stated to be bilinguals and thus were excluded from the experiment.

3.4.1 Correlation between L2 vocabulary size and knowledge of L2 collocations

The scores on the collocation task varied from 9 to 21 points ($M_{collocation\ task}$ = 15.7, *SD*= 2.75), 27 being the maximum. The scores on the vocabulary task ranged from 21 to 58 points ($M_{vocabulary\ task}$ = 39.4, *SD* = 9.62), 60 being the highest possible score. The mean and standard deviation are listed in Table 4; the scores on both tasks with percentage of correctness of each participant are depicted in Appendix 4.

***Table 4**: The mean and standard deviation for both tasks.*

	vocabulary task	collocation task
Mean	39.4	15.7
Standard deviation	9.62	2.75

In order to determine the strength of the relationship between knowledge of collocations and general vocabulary size, two variables, namely scores on both tasks, were tested and the correlation was calculated using an internet programme available at http://vassarstats.net/corr_stats.html. A Pearson product-moment correlation coefficient of r=.566 and $p < .001$ (N = 30) was found between the scores on the vocabulary task and the collocation task. This means that the scores on the vocabulary task explain R^2= 32% of the variance in the collocation task, that may be considered a rather small size effect. Based on the found correlation coefficient r= .566, it can be estimated there is a positive correlation between L2 vocabulary size and knowledge of L2 collocations; however, the strength of the relationship is only moderate. As illustrated in the scatter plot in Figure 3, the scores of the majority of the participants are to be found near the regression line.

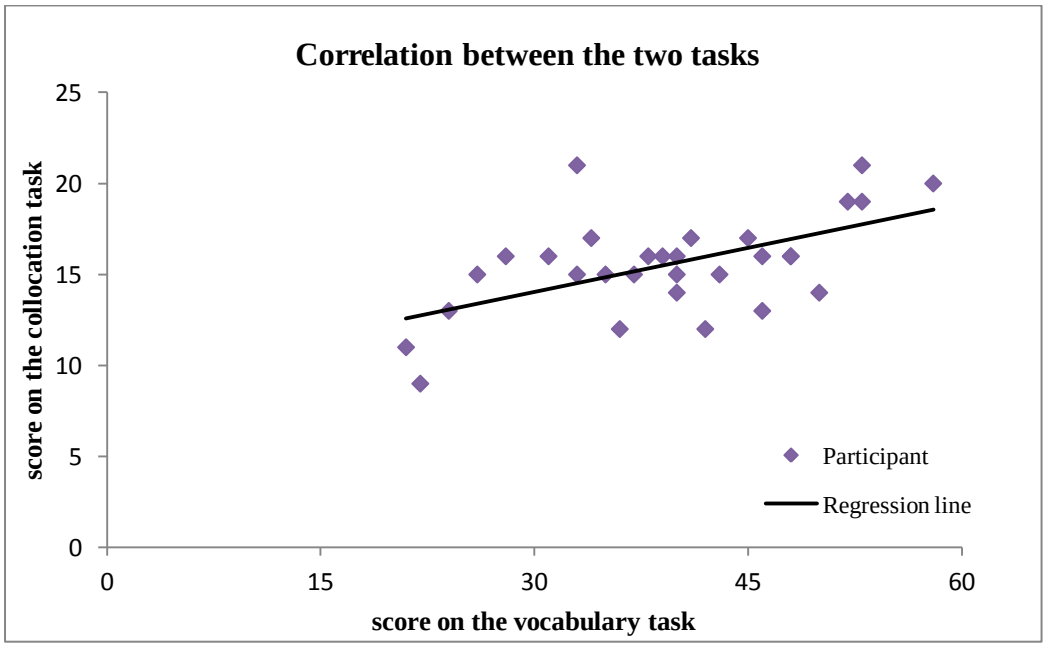

***Figure 3**: Scatter plot for the correlation between the scores on the vocabulary task and collocation task with r=.566, r^2=.32, slope =.162, intercept = 9.56 and standard error of estimate = 2.31.*

To sum up, a moderate positive correlation was observed between the scores on the vocabulary task and the collocation task. Moreover the participants' results do not vary considerably from the regression line. Nevertheless the results indicate the size of the L2 mental lexicon has a positive effect on knowledge of L2 collocations.

3.4.2 Knowledge of L2 collocations and similar patterns

Before analysing participants' knowledge of L2 collocations with similar patterns we looked at scores on the collocation task in more detail. In order to make the scale of tested L2 collocations broader, the collocation task comprised both verb-noun (VN) as well as adjective-noun (AN) L2 collocations. Although no predictions about the difference in the results of AN and VN collocations were included in our hypotheses, we can point out that scores on VN collocations were lower than on AN collocations (see Table 5). Out of 15 acceptable VN collocations used in the task the number of correctly identified collocations varied from 4 to 11 ($M_{VNcollocations}$ = 7.8 (52%), *SD* = 1.6). As already mentioned, participants' knowledge of AN collocations was slightly more profound. Out of 12 acceptable AN collocations used in the task, the amount of correctly identified collocations ranged from 4 to 10 ($M_{ANcollocations}$ = 7.8 (65%), *SD* = 1.8). Detailed scores of both VN collocations and AN collocations listed for each participant are to be found in Appendix 5.

Table 5*: The mean and standard deviation for verb-noun collocations and adjective-noun collocations.*

	VN collocations	AN collocations
Mean	7.8	7.8
Standard deviation	1.6	1.8

Furthermore, we can point out that participants' knowledge of English collocations does not necessarily go hand in hand with the frequency of usage of these collocations on the Internet. As illustrated in Table 6, although some collocations identified by the majority of the participants could be labelled as rather frequent, however, the most frequently used collocation in the collocation task *make effort* was recognized by only 70% of participants. In comparison to the collocation *give congratulations*, the least frequent VN collocation used in the task, it was correctly identified by 80% of participants. Moreover, it is interesting that while only 1 participant identified the collocation *pay respect,* its frequency is comparable with the collocation *give a compliment* that was recognized by 57% of participants.

Additionally, no collocation was recognized by all participants, as 2 participants failed to recognize collocations *heavy rain, pay attention* and *strong coffee*. Out of all provided collocations, *pay a compliment* is the only collocation not recognized by any participant. It has to be pointed out that no actual correlation between scores on the collocation task and frequency of occurrence was calculated. Since the frequency of occurrence of the tested L2 collocations on the Internet varied significantly and there is no limit according to which we can label collocations as frequent or infrequent, we decided not to use frequency as a variable. Therefore the above-mentioned data is based on a simple observation.

***Table 6**: Correctly identified collocations listed with frequency of occurrence on the Internet.*

English collocation	number of participants who correctly identified it (out of 30)	frequency
heavy rain	29	1 750 000
pay attention	29	17 200 000
strong coffee	29	394 000
dense forest	27	458 000
give thanks	26	3 010 000
lack patience	26	77 100
strong tea	26	409 000
thick pad	26	243 000
give congratulations	24	13 000
lose hope	24	1 230 000
wide influence	24	138 000
lose balance	23	217 000
thick eyebrow	23	21 400
make effort	21	162 000 000
lack courage	20	67 900
heavy snow	19	6 550 000
wide recognition	19	664 000
give a compliment	17	1 180 000
hard punch	16	278 000
lose temper	15	56 000
dense wood	14	173 000
lose patience	12	77 100
heavy eyebrow	3	4 200
give attention	2	435 000
pay respect	1	1 500 000
hold hope	1	109 000
pay a compliment	0	483 000

The second purpose of the experiment was to test whether already acquired L2 collocations can help learners in learning new collocations with similar patterns. As mentioned above out of 27 collocations provided in the collocation task, 8 VN and 8 AN collocations were selected for this purpose (as earlier depicted in Table 2). In order to analyze the data, an evaluation system was established. The participants were expected to recognize a frequent collocation (for instance *wide recognition*). If participants also recognized a rather infrequent collocation with a similar pattern (*wide influence*), they received a point. The maximum score was 9 points. In general about one-third of infrequent collocations selected for the purpose of this issue were recognized by the participants (see Table 7). With the exception of the collocations with the verb *give*, the participants were able to recognize more AN collocations with similar patterns than VN collocations. The mean and standard deviation are to be found in Table 8, a detailed list of scores for each participant is depicted in Appendix 6.

Table 7: *Number of participants (also percentage representation) who correctly identified also the infrequent collocations with similar patterns found in the frequent collocations.*

ADJECTIVE-NOUN collocations	Number of participants (out of 30)	Percentage of correct identifications	VERB-NOUN collocations	Number of participants (out of 30)	Percentage of correct identifications
Dense forest ***Dense*** wood	13	43	***Give*** *thanks* ***Give*** *a compliment*	15	50
Heavy snow ***Heavy*** rain	20	67	***Give*** *thanks* ***Give*** *congratulations*	21	70
Strong tea ***Strong*** coffee	23	77	***Lose*** *patience* ***Lose*** *temper*	4	13
Wide recognition ***Wide*** influence	10	33	***Pay*** *attention* ***Pay*** *respect*	1	3
			Pay *attention* ***Pay*** *a compliment*	0	0

Table 8: *The mean and standard deviation for the identification of collocations with similar patterns.*

	Collocations with similar patterns
Mean	3.4
Standard deviation	1.67

To summarize, the results indicate that participants' knowledge of L2 collocations are not inevitably influenced by the frequency of occurrence of these L2 collocations. In general approximately one-third of the participants were able to identify also rather infrequent collocations with similar patterns.

3.4.3 Correlation between L2 vocabulary size and interference of L1 on knowledge of L2 collocations

The third purpose of the experiment was to find out to what extent knowledge of L1 collocations influenced the participants when identifying L2 collocations in the collocation task. Out of all combinations of a verb and a noun or an adjective and a noun in the collocation task, 15 corresponded to Slovak collocations (8 VN collocations and 7 AN collocations as earlier depicted in Table 3). On average the participants incorrectly identified 5 such collocations as acceptable in English ($M_{L1errors}$ = 5.2, $SD_{L1errors}$ = 2.2). The number varied from 2 to 9, meaning every participant made at least two errors that can be attributed to transfer from Slovak. This indicates that no participant was able to identify only English collocations without being influenced by his/her L1 knowledge. If verb-noun and adjective-noun collocations are analyzed separately, it can be pointed out that the participants made slightly more mistakes that can be ascribed to interference on AN collocations ($M_{ANerrors}$= 2.7, $SD_{ANerrors}$ = 1.3) than on VN collocations ($M_{VNerrors}$ = 2.6, $SD_{VNerrors}$ = 1.4) (see Table 9). A detailed list of mistakes attributed to L1 transfer listed for each participant is depicted in Appendix 7.

***Table 9**: Mean and standard deviation of mistakes ascribed to L1 transfer.*

	Slovak collocations mistaken for English collocations	Slovak VN collocations mistaken for English collocations	Slovak AN collocations mistaken for English collocations
Mean	5.2	2.6	2.7
Standard deviation	2.2	1.4	1.3

The data was analyses of the correlation between L2 vocabulary size and number of errors attributed to interference in the collocation task. We found out there is indeed a positive correlation between the two variables, since found Pearson product-moment correlation coefficient was r =.26 and p= *NS* (N=30). Nonetheless the strength of the relationship between the tested variables is rather weak. The scatter plot in Figure 5 shows that only few

participants with high scores on the vocabulary task made fewer mistakes ascribed to L1 negative transfer. Moreover the results of the majority of the participants are to be found far away from the regression line, meaning the range of variation of the results is enormous and wide.

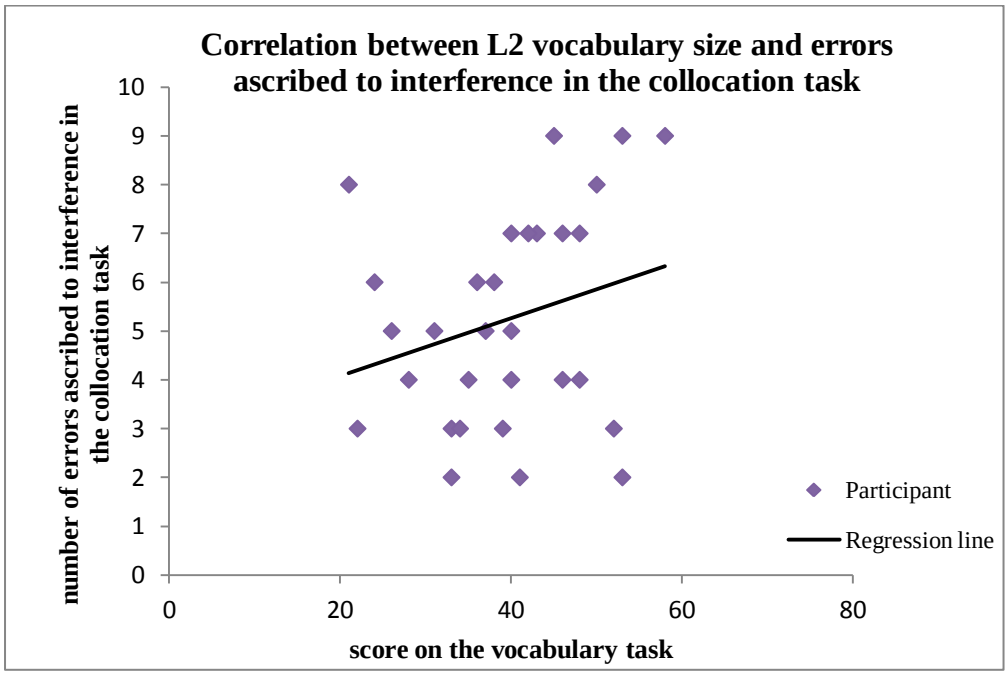

***Figure 5**: Scatter plot for the correlation between the scores on the vocabulary task and number of errors ascribed to interference in the collocation task with r=.26, r^2=.068, slope =.059, intercept =2.74 and standard error of estimate = 2.15.*

To summarize, a weak correlation was found between the scores on the vocabulary task and number of mistakes attributed to interference in the collocation task. However the relationship between the tested variables is rather weak. Moreover the participants' results varied considerably.

3.4.4 Correlation between frequency of occurrence of L1 collocations and interference of L1 on knowledge of L2 collocations

Our last research question centred on a possible positive correlation between frequency of occurrence of Slovak collocations and the amount of mistakes made in the collocation task attributed to L1 knowledge. As illustrated in Table 10, the incorrect identification of Slovak collocations as English collocations did not match the frequency of occurrence of these Slovak collocations on the Internet (as depicted in Table 3). Even though the most frequently

used Slovak collocation in the collocation task **heavy influence* was incorrectly identified as a correct English collocation by 18 participants, the majority of the participants identified collocations **hard operation* (*n* =26) and **make a compliment* (*n*=19) as possible in English although their frequency of occurrence on the Internet in Slovak was not so great. In addition, the data from Table 10 was further analysed in order to define the strength of the relationship between frequency of occurrence of 15 Slovak collocations and amount of mistakes made in the collocation task attributed to interference. A Pearson product-moment correlation of *r*=.144 and *p=NS* was found. However the frequency of the collocation **heavy influence* was too diverge from the other tested collocations, therefore it was evaluated as an outlier and excluded. A new correlation was calculated. A Pearson product-moment correlation of *r*= -.27 and *p=NS* was found. The relationship between the tested variables can be considered only moderate. A detailed scatter plot is to be found in Figure 6.

***Table 10**: Slovak collocations that were incorrectly identified as English collocations with frequency of occurrence on the Internet.*

Slovak collocation	number of participants who incorrectly identified it as English collocation (out of 30)	frequency
**hard operation*	26	3 390
**make a compliment*	19	839
**heavy influence*	18	121 000
**hold balance*	17	13 900
**strong punch*	14	11 200
**lose speech*	12	2 190
**hard wood*	11	41 300
**lose courage*	9	5 930
**lack temper*	8	5 780
**wide pad*	8	7 280
**lack hope*	5	705
**lack balance*	4	1 650
**give effort*	3	39 500
**thick snow*	2	31 200
**thick rain*	0	6 690

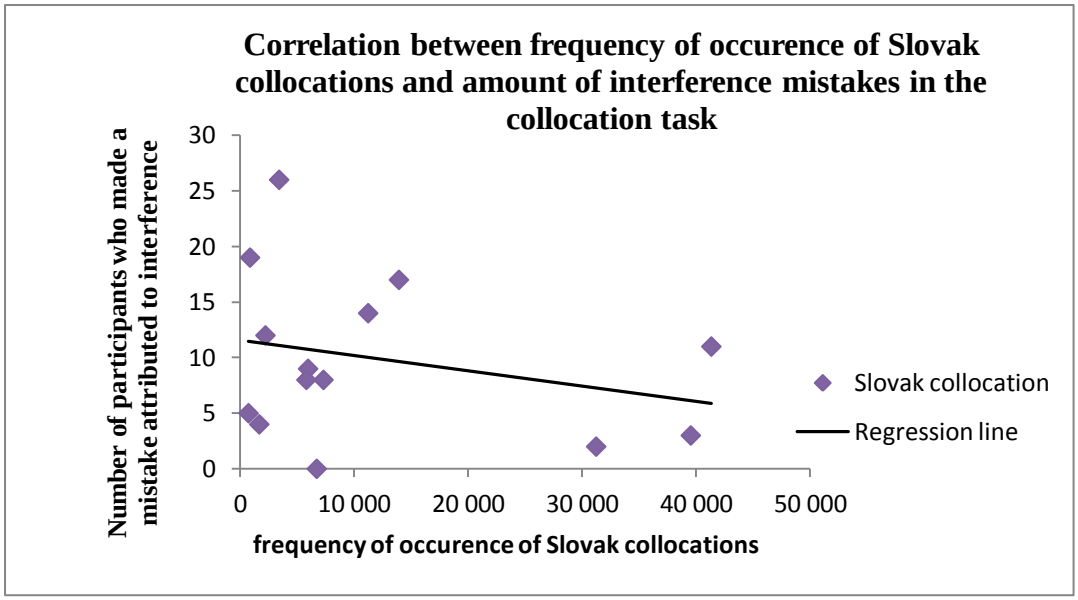

Figure 6: *Scatter plot for the correlation between the frequency of occurrence of Slovak collocations and number of made errors in the collocation task ascribed to interference with* $r = -.27$ $r^2 = .073$*, slope = 0, intercept =11.5 and standard error of estimate = 7.99.*

In conclusion, one collocation was excluded from the analysis since its frequency was too much far away from the rest. A moderate negative correlation was observed between the frequency of occurrence of Slovak collocations and number of mistakes in the collocations task that can be attributed to interference.

4 Discussion

4.1 Knowledge of L2 collocations and L2 vocabulary size

Our first hypothesis predicted that participants with higher scores on the vocabulary test will have better scores on collocation task than participants with lower scores on the vocabulary task. When discussing the correlation between knowledge of L2 collocations and L2 vocabulary size it is essential to firstly look at knowledge of L2 collocations more closely. As we found, it is interesting that the frequency of occurrence of the tested L2 collocations on the Internet did not play such a significant role as expected. In addition to identifying rather infrequent collocations, the majority of the participants also failed to recognize a few frequent collocations. Even the collocation *pay attention* that must very frequently occur in the input of students was not recognized by one participant. We may say that 97% correct identification of this collocation is very high and statistically significant; on the contrary we expected all participants to be familiar with such basic and frequently used collocation as *pay attention.* Moreover, due to the fact that collocations as *pay attention* do not have Slovak equivalents, meaning the concept is expressed differently in participants' L1, the participants may have recognized L2 collocations that do not occur in their L1 less effectively as in comparison to those that have L1 equivalents (Wolter & Gyllsad, 2011). However, as observed by Chan & Liou (2005), the assumption about a one-to-one correspondence may last longer in the case of infrequent collocations since the amount of input may determine the strength of the associations between concepts and L2 collocations. Nevertheless as stated in Shin (2006) collocations are more frequently used in spoken language than in written language. Therefore the amount of input in the spoken form may be the determining measure of frequency; however, our analysis was based on the occurrence of L2 collocations on the Internet in their written forms. On the other hand it would have been impossible to measure the actual frequency of the tested collocations in the input of the participants. Nonetheless we may suggest that the amount of input may not be the only determining factor. The length of collocations and their complexity may also have an impact on acquiring and recognizing such structures (Shin, 2006).

Based on the mean score on the collocation task being 15.6 (58%) we may conclude that the participants have gaps in their knowledge of L2 collocations. This conclusion is not in any sense surprising as previous studies on L2 collocations have shown that acquiring L2 collocations is exceedingly difficult for L2 learners (Bahns & Eldaw, 1993; Farghal & Obiedat, 1995; Gyllstad, 2007). One of the explanations why the participants did not gain

higher scores on the collocation task may be the fact that L2 mental lexicons are generally smaller than L1 mental lexicons, therefore an item in the L2 mental lexicon can only be directly connected to a small number of words creating word associations (Meara, 1996). In addition, L2 learners are unaware of the importance of L2 collocations in order to achieve mastery of L2 (Bahns & Eldaw, 1993) and they usually do not acquire L2 words with all competences comprised in word knowledge as proposed by Schmitt (1999). This may explain why the participants achieved higher scores on the vocabulary task measuring their general vocabulary than on the collocation task. We may claim only knowledge of written forms and words' meanings were necessary to correctly complete the vocabulary task; however, the collocation task required much more, namely knowledge of collocational behaviour and associations of acquired words. Thus we may conclude that our previous suggestion is indeed correct and L2 learners do not acquire words systematically with all their word competencies. Moreover, even though the amount of input plays an essential role in the acquisition process, it seems as if the results of the participants who often or very often read literature written in English and thus receive more native-like input did not vary widely from the results of participants who claimed never to read in English except for course books. In other words, the participants who never read did not gain fairly lower scores on both tasks as we might expect. Therefore we may suggest that in general even those participants who often read literature written in English do not pay much attention to such structures as L2 collocations and they are not aware of L2 collocations as a potential problem in SLA.

As the results depicted in the scatter plot in Figure 3 indicate, a positive correlation between knowledge of L2 collocations represented by scores on the collocation task and the size of the L2 mental lexicon represented by scores on the vocabulary task was found. The strength of the relationship between these two variables is only moderate; nonetheless we may claim the size of the L2 mental lexicon has a positive effect on the amount of knowledge of L2 collocations. Therefore we may conclude our first hypothesis to be borne out since the answer to the first research question is affirmative. Such conclusion was expected since L2 learners with bigger L2 vocabularies are generally more proficient in an extensive range of language skills (Meara, 1996), one of which being knowledge of L2 collocations. Other studies investigated a possible correlation between knowledge of L2 collocations and L2 vocabulary size as well (Bahns & Eldaw, 1993; Gyllstad, 2007). The results are rather contradicting, saying Bahns & Eldaw (1993) found no significant correlation between the two variables while Gyllstad (2007) observed rather high positive correlations ranging from .83 to .90. Even

though our results indicate a moderate correlation between the two variables, we may suggest the difference in the results might be explained with the differences in proficiency levels of participants since the above-mentioned studies investigated advanced L2 learners whereas we tested intermediate L2 learners with lower proficiency levels in L2 (B1 (n=11) or B2 level (n=19)). Bearing this in mind we may conclude there is indeed a positive correlation between knowledge of L2 collocations and general L2 vocabulary size; however, the proficiency level may have an impact on the strength of the relationship. Although CEFR[5] does mention collocations as being an essential part of learners' linguistic competences, it does not say anything about knowledge of what types of collocations or what amount of collocations is expected from learners of A1 – C2 proficiency levels. Therefore we may not with absolute certainty know whether the collocations used in the collocation task were adequate for the proficiency levels of the participants and whether its knowledge is to be expected from such L2 learners.

4.2 L2 collocations with similar patterns

In second hypothesis we predicted that participants with knowledge of frequently used L2 collocations will also be able to identify infrequent collocations with similar patterns as correct. In order for us to answer the second research question it is essential to look at the results with respect to this issue in more detail. Bearing the mean score of recognized collocations with similar patterns in mind (3.4) we can conclude that approximately one-third of the participants were able to identify also rather infrequent collocations with similar patterns. However we cannot be sure if the participants identified these collocations based on the analogy of similar patterns or whether it was just a coincidence that they identified both collocations of the "pair". Moreover it is essential to mention that the choice of tested L2 collocations may have influenced the results as well. We cannot with absolute certainty know whether the participants actually knew the frequent collocations based on which an analogy could be drawn. To put it differently, if the participants were not familiar with the collocation *dense forest,* we can presume they would not know the collocation *dense wood* as well, which would make it almost impossible for them to recognize the pattern. Due to the fact that there are no previous studies that investigated this issue, we cannot compare our results in order for us to better understand factors that influenced the performance of the participants.

[5] Retrieved from http://www.coe.int/t/dg4/linguistic/cadre1_en.asp

Furthermore, if we compare the participants who recognized many collocations with similar patterns in the collocation task with those who recognized only a few, we can suggest the participants' language background does not help us identifying the cause of the divergence of the results. A simple comparison of the results indicates that the participants with higher scores did not differ dramatically in their attitude towards using English or the amount of reading literature written in English from the participants with considerably lower scores. Overall, since the sample concerning knowledge of L2 collocations with similar patterns was quite small, the results cannot be considered statistically significant. We can conclude our hypothesis to be neither borne out nor rejected.

4.3 Interference and L2 vocabulary size

Our third hypothesis concerned interference, namely if mistakes in the collocation task ascribed to interference occur, participants with higher scores on the vocabulary test will make fewer such mistakes than participants with lower scores on the vocabulary test. The results contradict our hypothesis as we expected to find a negative correlation between scores on the vocabulary task and interference mistakes in the collocation task. Even though a positive correlation has been observed between the two variables, we may conclude the strength of the relationship is weak. Furthermore, the results of the participants vary considerably which does not make it possible for us to give a satisfactory answer on this issue, Nonetheless we may propose our hypothesis was not borne out. However, the experiment revealed various issues worth discussing. First of all it is interesting that quite a few participants who scored above average on the vocabulary task made many mistakes attributed to interference in the collocation task, which contradicts our earlier assumptions and generally acknowledged opinion that L2 learners with bigger L2 vocabularies are usually more proficient in an extensive range of language skills (Meara, 1996). We may assume that one of the explanations why the participants made such mistakes may be the fact that L2 learners are more easily influenced by their L1 knowledge if a tempting example is presented. In our case, out of all possible word combinations in the collocation task few responded to Slovak collocations that are incorrect in English. The participants might have been influenced by their L1 knowledge simply due to the fact that the strength of conceptual links between L1 and concepts are greater than between L2 and concepts. On the contrary, it does not necessarily mean that conceptual links between concepts and L2 are stronger for the participants who made only few such mistakes. We may suggest these participants might not

have been familiar with the English translations of the Slovak collocations, therefore they were less influenced by their L1 knowledge as other participants who made more mistakes attributed to interference. However, we can only suppose the mistakes made on the selected L2 collocations that corresponded to Slovak collocations are to be ascribed to interference. The participants' responses might have been evaluated just as errors. Moreover to label the tested collocations as incorrect in English may not be accurate. Although the correctness of L2 collocations was consulted with a dictionary, the "incorrect" collocations are to be found on Internet and are sometimes used even by native speakers. It is rather strict to define a word or an expression as ungrammatical and incorrect only because it has not yet been included in a dictionary.

The beginning stage of every L2 learner is L1 and throughout the process of SLA L2 learners gradually build their L2 lexical networks on the bases of input (Wolter, 2006). Although L2 lexical network slowly starts to diverge from the existing L1 network, L1 still has a strong influence on L2 lexical network formation (Wolter, 2006). Therefore we may claim that even if participants' responses were only mistakes, due to the fact that such expressions correspond to L1 collocations, interference must have influenced the participants' performance, even if subconsciously. Furthermore, the amount of input may determine the strength of associations between concepts and L2 collocations. The English translations of Slovak collocations used in the collocation task should not occur in the input of the participants, thus we may expect the participants to still base their performance on an incorrect assumption about a one-to-one correspondence between L1 and L2.

Another explanation why the participants with higher scores on the vocabulary task still made quite a few mistakes attributed to interference may be ascribed to paradigmatic and syntagmatic connections between words. Paradigmatic connections between L2 words are more easily established as the nature of the connections between words and concepts are not changed; L2 learners just need to substitute L1 words for the L2 words in their L2 mental lexicons. On the contrary syntagmatic connections are involved in conceptual modification (Wolter, 2006). Since syntagmatic connections creating word associations vary from language to language; L2 learners have to make new syntagmatic connections between individual words. Therefore we may claim gaining higher score on the vocabulary task, for which paradigmatic connections are necessary, is much easier than on the collocation task, where participants need syntagmatic connections. This means that the size dimension does not ensure that L2 production will not to be influenced by L1 knowledge. This brings us back to

the previously mentioned suggestion, saying that not the size but the structure is the essential feature of the L2 mental lexicon.

It is very difficult to compare our results to previous studies investigating L2 collocations and interference (Farghal & Obiedat, 1995; Gabrys-Biskup, 1992; Gyllstad, 2007; Wolter & Gyllstad, 2011), as none of them focused on the correlation between such mistakes and L2 vocabulary size. Moreover the results of the only study with a similarly structured task (Gyllstad, 2007) indicate that low proficiency learners are more likely to rely on their L1 knowledge and thus make more mistakes ascribed to negative transfer. Since our participants did not vary in proficiency levels, we cannot compare the results. Nonetheless we may suggest the participants' proficiency and not general L2 vocabulary size might have been the determining factor for the strength of the impact of interference on identifying L2 collocations. We may conclude the number of mistakes ascribed to L1 transfer made in the collocation task does not necessarily go hand in hand with the general L2 vocabulary size. It seems as if the proficiency may have a bigger influence on interference with respect to L2 collocations. However, we can only speculate whether such assumptions are true as our results vary considerably.

4.4 Interference and frequency

Our final hypothesis concerned correlation between interference mistakes and frequency, namely if mistakes in the collocations task ascribed to interference occur, the participants would be more influenced by their knowledge of rather frequent Slovak collocations than infrequent ones. After excluding the collocation **heavy influence* from the analysis, a negative correlation was found (-.27) between the two variables. However, we cannot consider it to be statistically significant as the tested sample comprised only 14 collocations which do not allow us to draw a definite and valid conclusion. Despite the limitations we can suggest that participants were not influenced by knowledge of frequently used Slovak collocations as expected. We may claim that although our hypothesis was not borne out, the idea need not necessarily be rejected. It seems very logical that L2 learners would be more influenced by their knowledge of frequently used L1 collocations. Nevertheless the measure of frequency may not have been very accurate. Firstly frequent usage of L1 collocations on the Internet does not mean that participants also use them in everyday speech. Secondly to label collocations

as frequent or infrequent is rather subjective, vague and context dependent since for instance the collocation **hard wood* can be judged as frequent in comparison to the collocation **make a compliment* but as infrequent in comparison with the collocation **heavy influence.* We have to mention that the tested collocations are also used on the Internet in English forms, which means that L2 learners could also be confronted with such expressions in L2. Moreover the independency of the results on the frequency can be accounted for other factors. The cause why many participants identified infrequent collocation **make a compliment* as correct in English may be explained by the fact that at the time of testing they might have been studying word combinations in English with the verb *make,* therefore they overgeneralized the usage of this verb.

4.5 Conclusion

After testing a group of secondary school students studying English in Slovakia and their knowledge of L2 collocations we have found a positive correlation (.566) between general L2 vocabulary size and knowledge of L2 collocations. This conclusion was expected since L2 learners with bigger L2 vocabularies are usually more proficient in various language skills. Nonetheless we suggested some explanations why the participants gained higher scores on the vocabulary task than on the collocations task. We believed the structure of the L2 mental lexicon could be identified as the main cause since it seemed as if participants' associations between words were not as strong as in comparison to associations between words with respect to the L1 mental lexicon. To put it differently, to gain high scores on the vocabulary task where only knowledge of individual words is necessary is easier than to achieve high scores on the collocation task for which knowledge of word associations and their collocational behaviour is vital.

In our second research question we asked whether participants would identify infrequently used L2 collocations with similar patterns found in the frequent L2 collocations. Due to the fact that no previous studies investigated this issue we were unable to compare our results in order to try to better understand them and find a possible cause for the variance of the results. Since the sample of tested L2 collocations with respect to similar patterns was small, we concluded it might be difficult to define whether the participants made use of the analogy or if it was a simple coincidence. In addition, as scores on the collocation task concerning this issue varied widely, we tried to attribute the variance to attitude toward using L2 or reading

literature written in L2. However, no major difference in the results was found as participants with positive attitude toward using L2 or reading in L2 generally did not gain higher scores on the collocation task.

Our third hypothesis was focused on negative transfer. Our results indicated a positive correlation (.309) between general L2 vocabulary size and amount of mistakes made in the collocation task attributed to interference; nevertheless only a rather weak correlation was observed. We concluded the L1 still plays a significant role in the SLA of the participants as the associations between concepts and L1 are generally stronger and thus preferred over the associations between concepts and L2. Therefore L1 is used as a mediator which makes it easier to be influenced by L1 knowledge, especially in the earlier stages of SLA. Since even participants with higher scores on the vocabulary task did not make less such mistakes as previously expected we concluded the proficiency level of the participants was not profound enough in order for interference not to play such a significant role.

Interestingly the frequency of occurrence of tested L2 collocations on the Internet did not play such a significant role in participants' scores as expected. Although a negative correlation (-.27) was found between frequency of occurrence of L1 collocations on the Internet and number of participants who identified these collocations as correct in L2, we concluded the tested sample was too small in order for us to draw a definite and valid conclusion. We suggested that to test the actual frequency of occurrence of tested collocation in the input of the participants would have been better for the purposes of the experiment; however, we concluded it would have been impossible to measure the occurrence of tested collocations in the input.

4.6 Suggestions for further research

Due to the fact that our results and discussion indicate there are some issues that need to be examined in more detail, we propose several suggestions for further research. Firstly we think further research is necessary that would explain the gap in knowledge of L2 collocations occurring in the input very frequently given the gap cannot be attributed to interference (as in the example of L2 collocation *pay attention*). Secondly a detailed research on the acquisition of collocations should give better answers on the question whether L2 learners actually make use of the analogy to facilitate the acquisition process of L2 collocations. Last but not least it has been generally acknowledged that reading in L2 is beneficial for L2 learners who in this

way receive more native-like input and can improve in various language skills as grammar or vocabulary. Our results indicate that, in general, the participants who read very often did not get higher scores on the collocation task. Thus we believe further research may reveal whether reading is beneficial for acquiring L2 collocations and what type of written texts can be considered the best input with respect to L2 collocations.

Work cited

BAHNS, J., & ELDAW, M. (1993). Should we teach EFL students collocations. *System*, 21, 101-114.

BARFIELD, A. (2003). *Collocation Recognition and Production: Research Insights*. Tokyo: Chuo University.

BONK, W.J. (2000). Testing ESL learners' knowledge of collocations. Educational Resources Information Centre. Research Report ED 442 309.

CHAN, T. & LIOU, H. (2005), Effects of Web-based Concordancing Instruction on EFL Students' Learning of Verb – Noun Collocations, *Computer Assisted Language Learning*, 18, 231 – 250.

COWAN, L. (1989). *Towards a definition of collocation*. Unpublished master's thesis. Concordia University, Montreal.

DÖRNYEI, Z. (2003). Attitudes, orientations, and motivations in language learning: Advances in theory, research, and applications. *Language Learning*, *53*(S1), 3-32.

DULAY, H.C. & BURT, M.K. (1973). Should we teach children syntax? *Language Learning*, 23, 245-258.

ELLIS, R. (1994). *The Study of Second Language Acquisition*, Oxford: Oxford University Press.

ELLIS. N. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition* 18, 91-126.

FARGHAL, M., & OBIEDAT, H. (1995). Collocations: A neglected variable in EFL. *International Review of Applied Linguistics,* 33, 315-331.

GABRYS-BISKUP, D. (1992). L1 influence on learners' renderings of English collocations: A Polish/German empirical study. In P. J. L. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (85-93). London: Macmillan.

GITSAKI, C. (1996). *English collocations in second language acquisition.* Unpublished Ph.D. dissertation. University of Queensland. Australia.

GLASER, R. (1998). The grading of idiomaticity as a presupposition for a taxonomy of idioms. In Hüllen, W., & Schulze, R. (Eds.). (1988). *Understanding the lexicon: meaning, sense and world knowledge in lexical semantics* (Vol. 210). Walter de Gruyter.

GYLLSTAD, H. (2007). *Testing English Collocations. Developing Receptive Tests for Use with Advanced Swedish Learners*. Unpublished Ph.D. dissertation. Lund University, Sweden.

HA, M. (1988). Acquisition of collocations: Investigation of relative difficulty among different types of collocations. Unpublished manuscript, University of Hawaii.

HOWARTH, P. (1998). Phraseology and Second Language Proficiency. *Applied Linguistics* 19(1). 24-44.

HUANG, L.-S. (2001). Knowledge of English Collocations: An Analysis of Taiwanese EFL Learners. Resources Information Centre. Research Report ED 465 288.

HUANG, X. Y. (2011). A study on the relationship between university students' Chinese writing proficiency and their English writing proficiency. *English Language Teaching*, 4, 55-65.

JERVIC, T (2011). The Study of the Problem of Absolute Syonyms in Relation to their Collocational Behaviour. *English Language and Anglophone Literatures Today (ELALT)*. 193-199.

KRASHEN, S.D. (1981), *Second Language Acquisition and Second Language Learning*, Unversity of California: Pergamon Press Inc

KROLL, J. F., & STEWART, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of memory and language*, *33*, 2, 149-174.

KROLL, J. F. et al. (2002). The development of lexical fluency in a second language. *Second language research*, 18, 2, 137-171

LADO, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.

LAUFER, B. & NATION, P. (1999). A vocabulary-size test of controlled productive ability, *Language Testing*, 16, 33–51

LIN, D. (1999) Automatic identification of non-compositional phrases. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics* . Association for Computational Linguistics, 317-324.

MEARA, P. (1996). The dimensions of lexical competence. In Brown, G., Malmkjaer, K. and Williams, J. (eds.), *Performance and Competence in Second Language Acquisition* (35-53). Cambridge: Cambridge University Press

MOCHIZUKI, M. 2002. Exploration of two aspects of vocabulary knowledge: Paradigmatic and collocational. *Annual Review of English Language Education in Japan* 13. 121-129.

MOYER, A. (1999). Ultimate attainment in L2 phonology: the critical factors of age, motivation and instruction. *Studies in Second Language Acquisition*, 21, 81-108.

NATION, I.S.P. (2001), *Vocabulary size test - Learning Vocabulary in Another Language*, Cambridge: Cambridge University Press

RICHARDS, J. C.& SCHMIDT, R. (2002). *Longman Dictionary of Language Teaching and Applied Linguistics (3rd ed.).* London: Pearson Education Limited.

SCHMITT, N. (1999). The relationship between TOEFL vocabulary items and meaning, association, collocation and word-class knowledge, *Language Testing,* 16, 189–216

SCHMITT, N., SCHMITT, D. & CLAPHAM, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test, *Language Testing* ,18, 55–88

SHIN, D. (2006). *A collocation inventory for beginners*. Unpublished Ph.D. dissertation. Victoria University of Wellington, New Zealand.

SIEPMANN, D. (2005). Collocation, colligation and encoding dictionaries. Part I: Lexicological Aspects. *International Journal of Lexicography*, *18*(4), 409-443.

SRIDHAR, S. N. (1975). Contrastive Analysis, Error Analysis and Interlanguage: Three Phases of One Goal?, *Studies in Language Learning*, Vol. 1.

TRAN-CHI-CHAU. (1975). Error and explanations. *Applied Linguistics*, 7, 144-166.

UNSWORTH, S. (2008). Age and input in the acquisition of grammatical gender in Dutch. *Second Language Research* 24: 365-396

UNSWORTH, S. & BLOM, E. (2010). Comparing L1 children, L2 children and L2 adults. *Experimental methods in language acquisition research*, 201-222.

WOLTER, B. (2001). Comparing the L1 and L2 mental lexicon. *Studies in Second Language Acquisition, 23,* 01, 41-69.

WOLTER, B. (2006). Lexical Network Structures and L2 Vocabulary Acquisition: The Role of L1 Lexical/Conceptual Knowledge. *Applied Linguistics*, 27, 741 – 747.

WOLTER, B. & GYLLSTAD, H. (2011). Collocational Links in the L2 Mental Lexicon and the Influence of L1 Intralexical Knowledge, *Applied Linguistics*, 32, 430–449.

WRAY, A. (2002), *Formulaic language and the lexicon.* Cambridge UK: Cambridge University Press.

ZAREVA, A. (2007), Structure of the second language mental lexicon: how does it compare to native speakers' lexical organization?, *Second Language Research* 23, 2, 123–153.

Appendices

Appendix 1: A detailed table of the participants

ID	gender	age	age of onset	years of exposure	English lessons per week	Proficiency in English	attitude to using English	amount of reading in English	time spent abroad	knowledge of other foreign languages
1	m	17	6-10	11	4	B2	a lot	never		2
2	m	18	-6	13	+5	B2	very much	very often		1
3	m	19	6-10	13	+5	B1	neutral	often		2
4	m	18	6-10	9	4	B1	neutral	sometimes		2
5	m	18	6-10	10	+5	B2	a lot	sometimes		1
6	f	18	6-10	11	4	B1	neutral	sometimes		0
7	m	17	6-10	10	+5	B2	not really	sometimes		2
8	m	18	11-15	5	5	B2	neutral	sometimes		2
9	f	17	6-10	10	4	B2	neutral	sometimes		1
10	m	19	11-15	4	+5	B2	very much	very often	2 months	1
11	m	18	6-10	12	+5	B2	very much	often		1
12	m	18	6-10	8	5	B1	very much	sometimes		1
13	m	18	-6	15	4	B2	neutral	sometimes	2 month	1
14	m	18	6-10	9	4	B2	very much	sometimes	2 weeks	2
15	m	18	11-15	7	4	B1	a lot	sometimes	2 weeks	2
16	m	18	6-10	9	4	B1	neutral	sometimes		1
17	m	18	11-15	7	5	B2	a lot	often	2 weeks	2
18	f	18	11-15	9	4	B2	a lot	often		2
19	m	18	-6	13	+5	B2	a lot	often		1
20	f	18	-6	13	+5	B2	very much	often		2
21	f	17	11-15	6	4	B2	not really	sometimes		2
22	m	18	11-15	8	4	B2	neutral	never		2
23	f	18	6-10	9	4	B1	a lot	sometimes		1
24	f	17	-6	13	+5	B1	not really	never		1
25	f	17	-6	12	5	B1	a lot	sometimes		0
26	f	18	-6	4	4	B2	a lot	sometimes		1
27	m	17	6-10	9	5	B1	a lot	sometimes		2
28	m	18	11-15	4	4	B1	neutral	sometimes		0
29	m	18	6-10	10	+5	B2	not really	sometimes		0
30	m	18	6-10	9	4	B2	a lot	often		2

Appendix 2: The collocation task as presented to the participants.

There are 4 tables. In every table there are 3 verbs or 3 adjectives on the left and 6 nouns on the top. Your task is to put a cross (X) in a cell where two words (a verb and a noun or an adjective and a noun) come together, creating a good pair (as depicted in the example). A word on the left can be combined with more than one noun. If you decide to change your answer, please specify your final answer under the corresponding table. You have 8 minutes to finish this task.

Example

	a challenge	*damage*	*justice*	*a murder*	*a problem*	*suicide*
commit				*X*		*X*
do		*X*	*X*			
solve	*X*			*X*	*X*	

1.

	balance	courage	hope	patience	a speech	temper
hold						
lack						
lose						

2.

	attention	a compliment	congratulations	an effort	respect	thanks
give						
make						
pay						

3.

	eyebrow	influence	pad	rain	recognition	snow
heavy						
thick						
wide						

4.

	coffee	forest	operation	punch	tea	wood
dense						
hard						
strong						

Appendix 3: Sociolinguistic questionnaire as presented to the participants

1. What is your gender?
    a) Male
    b) Female

2. How old are you?
    _____ years old

3. What is your mother language?
    a) Slovak
    b) Other

4. Were you raised a bilingual (do you have two mother languages)?
    a) No
    b) *Yes
5. If you answered "Yes" to the previous question, please specify the languages:
    _____

6. How many foreign languages other than English do you speak?
    a) 0
    b) 1
    c) 2
    d) More than 2

7. When did you first start learning English?
    a) Younger than 6 years old
    b) Aged 6 – 10
    c) Aged 11 – 15
    d) Aged 16 – 17

8. How many years have you been learning English?
    _____ years

9. How many lessons of English do you have per week?
    a) 1 – 2
    b) 3
    c) 4
    d) 5
    e) More than 5

10. Have you ever attended any additional English courses/lessons outside of school?
    a) No, never
    b) I used to, but I do not anymore
    c) Yes, right now I am attending additional English courses/lessons

11. If you answered b) or c) in the previous question, please define for how long you have been attending such courses/lessons (e.g. *two lessons per week for three years*).
    _____

12. What is your proficiency in English?
    a) Beginner (A1)
    b) Pre-Intermediate (A2)
    c) Intermediate (B1)
    d) Upper-Intermediate (B2)
    e) Advanced (C1)

13. Do you have any certificates of your English proficiency (TOEFL, IELTS, FCE, CAE,..)?
    a) No
    b) *Yes

14. If you answered "Yes" to the previous question please specify the name of the certificate, when did you get it and what was the proficiency level/overall result (e.g. *IELTS, June 2012, overall result 7,5*).
    _____

15. How often do you read in English (excluding reading the course book)?
    a) Never
    b) Sometimes
    c) Often
    d) Very often

16. Have you spend any time in an English speaking country?
    a) No
    b) Yes

17. If you answered "Yes" to the previous question, please define the country/city of stay, for how long have you been there and the purpose of the stay (e.g. *London, 2 months, attending English classes*).
    _____

18. How much do you like using English?
    a) Very much
    b) Quite a lot
    c) Neutral
    d) Not really
    e) Not at all

Appendix 4: Scores on both tasks with percentage of correctness listed for each participant.

ID	Scores on the collocation task (out of 27)	Percentage of correct answers on the collocation task	Scores on the vocabulary task (out of 60)	Percentage of correct answers on the vocabulary task
1	16	59.3	48	80
2	20	74.1	58	96.7
3	14	51.9	40	66.7
4	21	77.8	33	55
5	15	55.6	43	71.7
6	16	59.3	39	65
7	16	59.3	48	80
8	16	59.3	40	66.7
9	11	40.7	21	35
10	16	59.3	31	51.7
11	16	59.3	46	76.7
12	19	70.4	53	83.3
13	13	48.2	46	76.7
14	17	63	45	75
15	15	55.6	40	66.7
16	13	48.2	24	40
17	19	70.4	52	86.7
18	21	77.8	53	83.3
19	16	59.3	38	63.3
20	17	63	41	68.3
21	15	55.6	35	58.3
22	12	44.4	42	70
23	15	55.6	33	55
24	15	55.6	37	61.7
25	16	59.3	28	46.7
26	17	63	34	56.7
27	12	44.4	36	60
28	9	33.3	22	36.7
29	15	55.6	26	43.3
30	14	51.9	50	83.3

Appendix 5: Detailed scores of verb-noun collocations and adjective-noun collocations with percentage of correct answers listed for each participant.

ID	Number of correct collocations (out of 27)	Number of correct VN collocations (out of 15)	Percentage of correct VN collocations	Number of correct AN collocations (out of 12)	Percentage of correct AN collocations
1	16	9	60	7	58.3
2	20	10	66.7	10	83.3
3	14	7	46.7	7	58.3
4	21	11	73.3	10	83.3
5	15	6	40	9	75
6	16	9	60	7	58.3
7	16	6	40	10	83.3
8	16	6	40	10	83.3
9	11	4	26.7	7	58.3
10	16	9	60	7	58.3
11	16	9	60	7	58.3
12	19	9	60	10	83.3
13	13	7	46.7	6	50
14	17	8	53.3	9	75
15	15	8	53.3	7	58.3
16	13	7	46.7	6	50
17	19	8	53.3	11	91.7
18	21	10	66.7	11	91.7
19	16	8	53.3	8	66.7
20	17	9	60	8	66.7
21	15	8	53.3	7	58.3
22	12	8	53.3	4	33.3
23	15	7	46.7	8	66.7
24	15	7	46.7	8	66.7
25	16	8	53.3	8	66.7
26	17	8	53.3	9	75
27	12	6	40	6	50
28	9	5	33.3	4	33.3
29	15	7	46.7	8	66.7
30	14	8	53.3	6	50

Appendix 6: A list of scores on the collocation task concerning knowledge of collocations with similar patterns listed for each participant.

ID	Score on the collocation task concerning collocations with similar patterns (out of 9)
1	4
2	6
3	4
4	6
5	2
6	5
7	5
8	4
9	3
10	2
11	3
12	4
13	1
14	3
15	2
16	2
17	5
18	7
19	3
20	5
21	2
22	0
23	3
24	3
25	5
26	4
27	4
28	1
29	3
30	1

Appendix 7: A detailed list of mistakes attributed to transfer from Slovak for verb-noun collocations and adjective-noun collocations listed for each participant.

ID	Number of mistakes ascribed to L1 transfer (out of 15)	Number of mistakes of VN collocations ascribed to L1 transfer (out of 8)	Number of mistakes of AN collocations ascribed to L1 transfer (out of 7)
1	4	1	3
2	9	3	5
3	7	4	3
4	3	2	1
5	7	3	4
6	3	2	1
7	7	4	3
8	5	3	2
9	8	5	3
10	5	1	4
11	7	4	3
12	9	5	4
13	4	1	3
14	9	5	4
15	4	2	2
16	6	2	4
17	3	2	1
18	2	1	1
19	6	3	3
20	2	2	0
21	5	1	4
22	7	3	4
23	2	1	1
24	5	3	2
25	4	2	2
26	3	2	1
27	6	3	3
28	3	1	2
29	5	3	2
30	8	5	5