Factors Impacting Recognition of False Collocations by Speakers of English as L1 and L2

Olga Makinina

Currently there is a general uncertainty about what makes collocations (i.e., fixed word combinations with specific, not easily interpreted relations between their components) hard for ESL learners to master, and about how to improve collocation recognition and learning process. This study explored and designed a comparative classification of external factors that affect adult English as L1 and L2 speakers' recognition of false (non-native-like) collocations. At Stage 1, the reading-comprehension test and interview were administered to 5 participants: emergent bilinguals (another language/English), an advanced bilingual (Russian/ English), and a monolingual speaker of English. The results provided insights into the strategies and criteria participants employed to identify false collocations. At Stage 2, more than 90 speakers of English as L1 and L2 took a reading-comprehension test and posttest survey that gathered information on the differences and similarities in the participants' collocation recognition. The results suggested that certain factors positively correlated with recognition of false collocations: English as a predominant language of communication, the length of residence, vocabulary-learning strategies, and the focus of attention on sentence structure and the form and meaning of word combinations. The implications concern potential focus areas for pedagogical intervention in the ESL classroom.

Une incertitude généralisée plane quant à la raison pour laquelle la maitrise des expressions figées (c.-à-d., des combinaisons de mots unis par des liens spécifiques et difficiles à interpréter) est problématique pour les apprenants d'ALS, et quant aux moyens d'améliorer la reconnaissance et l'apprentissage des expressions figées. Nous avons conçu et ensuite exploré une classification comparative des facteurs externes qui affectent la reconnaissance, par des adultes locuteurs natifs d'anglais et des adultes dont l'anglais est la langue seconde, de fausses expressions figées. La première étape a consisté en un test de compréhension à l'écrit et une entrevue avec 5 participants: des bilingues émergents (autre langue/anglais), un bilingue avancé (russe/anglais) et un locuteur unilingue d'anglais. Les résultats offrent des informations sur les stratégies et les critères qu'emploient les participants pour identifier de fausses expressions figées. Pendant la deuxième étape, plus de 90 locuteurs natifs d'anglais et de locuteurs dont l'anglais est la langue seconde ont complété un examen de compréhension de lecture et une enquête de suivi portant sur les différences et les similarités dans la reconnaissance par les participants des expressions figées. Les résultats laissent entrevoir une corrélation positive entre certains facteurs et la reconnaissance de fausses expressions figées: le fait d'avoir l'anglais comme langue dominante, la durée du séjour, les stratégies

pour apprendre le vocabulaire, et l'attention portée à la structure des phrases ainsi qu'à la forme et au sens des combinaisons de mots. Cette recherche a des retombées quant aux domaines potentiels sur lesquels pourraient porter les interventions pédagogiques dans les cours d'anglais langue seconde.

Although in the past few decades, the focus of second language acquisition (SLA) research on the lexical proficiency in English as L2 has shifted from individual words as basic lexical units to formulaic word sequences—in particular, collocations and factors that influence their acquisition as explored in the studies by Gitsaki (1999), Nesselhauf (2005), Nguyen and Webb (2016), Pellicer-Sánchez (2015), Wolter and Gyllstad (2013), and Ying and O'Neill (2009)—there is still no clear understanding of what factors contribute to or impede collocation acquisition and consequently might predict the conditions of success in collocation learning and teaching. In part, this is due to the ambiguous nature and the absence of a universally agreed-upon definition or typology of collocations. The working definition of collocations focuses on the fact that these fixed multiword combinations penetrate a language and might belong to different genres, thus becoming a complex code system used by speakers of this language to render meaning in an efficient and comprehensible manner (Firth, 1957; Gitsaki, 1996; Schmitt, 2010; Wray, 2002). However, general inexplicability of relations between collocation components leads some researchers such as Siyanova and Schmitt (2008) and Siyanova-Chanturia and Spina (2015) to a claim that collocation knowledge is largely based on intuition that L2 learners might not possess. Thus, collocation learning is a slow process, and often even advanced L2 learners are unable to master these word combinations successfully (Groom, 2009; Nesselhauf, 2005).

Several of those studies that examined collocation acquisition (Cieślicka, 2015; Gyllstad, 2009; Henriksen, 2013) identified that speakers of English as L2 have difficulty with not only producing but also recognizing collocations. Specifically, they might not be able to differentiate collocations from nonnative-like word combinations, that is, miscollocations or false collocations (Barnbrook, Mason, & Krishnamurthy, 2013; Wray & Perkins, 2000). This means that, unlike speakers of English as L1 who process formulaic units holistically, ESL learners process/recognize word combinations analytically, that is, one-by-one, and treat the co-occurrence of the words as random. This causes a large number of collocational errors that range from direct translation of collocations from L1 to L2 to their paraphrasing, that is, substituting words for their synonyms, or blending as discussed in the studies by Cieślicka (2015), Henriksen and Stoehr (2009), and Wray (2002).

OLGA MAKININA

This finding of the previous studies motivated the focus of the current study on recognition of false (non-native-like) collocations by speakers of English as L1 and L2 and those factors that might impact this recognition. The majority of previous studies address only one or several factors, for example, the predominant language for thinking and communication (Wang & Shih, 2011), the frequency and quality of the language input (Cieślicka, 2015; Szudarski & Carter, 2014), the age of a learner (Granena & Long, 2013; Wray, 2002), the absence of the native speakers' intuition (Gitsaki, 1996; Siyanova & Schmitt, 2008), the language proficiency level (Groom, 2009; Li & Schmitt, 2010), and L1-L2 inter-influence (Bylund, Abrahamsson, & Hyltenstam, 2012; Liao, 2010; Millar, 2011). However, to the best of my knowledge, there have been no attempts to create a comprehensive classification of multiple factors. This study aims at addressing this issue by identifying patterns and correlating these factors across the groups of speakers of English as L1 and L2.

Factors Influencing Collocation Recognition

Although there is no generally accepted classification of the factors that impact collocation recognition, the factors addressed in this study and frequently mentioned in different research works were combined into a multileveled typology for the purpose of this research. It is necessary to differentiate between interlinguistic factors, such as English language proficiency, the predominant language of communication and thinking, L1-L2 inter-influence, and code-switching; extralinguistic factors, such as the age of onset of learning English, the length of residence in an English-speaking country, and the amount and quality of the language input; and cognitive factors, such as vocabulary learning strategies, reading fluency, and criteria of choosing incorrect collocations (the focus on the word form, meaning, both, or something else).

Interlinguistic Factors

Most studies focus on the English language proficiency level as related to collocational errors. However, the researchers' opinions differ slightly regarding the errors that characterize proficiency levels. Thus, Groom (2009) and Li and Schmitt (2010) argue that at a lower proficiency level, learners tend to overuse formulaic units, combining their components incorrectly or paraphrasing. On the contrary, Wray (2002) notes that beginning learners use collocations-clichés, that is, highly frequent and recurrent routine formulas used in habitual communicative situations, such as *Good morning* (Krishnamurthy, 2005), and only later, when they reach the intermediate level, start combining words creatively, that is, substituting collocation components with their synonyms. That generates the "accuracy of use" errors (Schmitt, 2010) and consequently might trigger production of false collocations. While discussing this type of error, some researchers explain that it is directly impacted by using L1 or English for inner speech. Han (2004) mentions L1-based thinking that is re-

flected in the L2 speaking and/or writing and "L1 preprogramming" that might decrease the learners' sensitivity to L2 formulaic units. Wang and Shih (2011) argue that using English as a predominant language of communication and thinking will decrease direct translation or paraphrasing, that is, production of false collocations.

Further expanding upon the idea of L1 influencing collocational errors, the research indicates that certain types of collocations might be more difficult for recognition depending on the learners' primary language because L1-L2 inter-influence might take place. Bylund et al. (2012), MacWhinney (2006), and Nakuma (1998) discuss different forms of L1 and L2 interaction, such as competition, parasitism of L2 that involves L2 development as dependent on L1, L1 resonance in the structures and semantics of L2, and L1 entrenchment (automatization). These factors make second language learners paraphrase or blend collocations, or implement direct translation from L1 to L2. Millar (2011) and Zyzik (2009) discuss similarity of multiword units in L1 and L2 as a significant factor influencing difficulty of comprehension. Conversely, Wolter and Gyllstad (2011) underline differences between L1 and L2 structures and claim that the main reason for difficulty in collocation recognition is a greater level of variability of hierarchical structures that combine words across languages. They argue that connections between words in collocations are dependent on a particular language (e.g., *plastic surgery* in English and пластическая операция [plastic operation] in Russian), and therefore are more difficult for language learners to recognize. Jafarpour, Hashemian, and Alipour (2013) and Webb and Kagimoto (2011) focus on synonyms for collocates and nodes that might be similar or different in L1 and L2, which again impedes with collocation comprehension. Additionally, Peters (2016) identifies that although congruent collocations are less likely to be problematic for recognition/recall in comparison to incongruent formulaic units, both are still a source of challenge for learners.

Extralinguistic Factors

Researchers, for example, Nesselhauf (2003), who focused on advanced learners, and Levitzky-Aviad and Laufer (2013), who studied the ESL learners' progress for 8 years, agree on the fact that collocational errors are ineradicable until the near-native proficiency level is attained. However, the amount of time it takes learners to achieve this level of collocation proficiency might differ depending on the age of onset of learning English. When comparing collocational errors across second language learners, Granena and Long (2013) identified that the ability to recognize and acquire collocations is at its best from 0 to 6 years old, after which it starts decreasing. Wray (2002) claims that these differences lie in the discrepancies between the adults' and children's learning strategies. Unlike children, who process word combinations holistically, adults engage in the analysis of the multiword units, break them down, and memorize their components as individual words. Later on, they

might overlook the initial links between collocates and their nodes and start interchanging words, namely substituting collocates into their synonyms (Jafarpour et al., 2013) which makes them sound non-native-like. The length of residence in an English-speaking country and the frequency and quality of the input are the factors promoted in the studies that discuss the benefits of the implicit learning approach and language immersion (Alali & Schmitt, 2012; Durrant & Schmitt, 2010) in terms of developing collocational fluency. The researchers argue that frequent exposure and practice might increase the learners' intuition and natural collocation awareness. At the same time, Szudarski and Carter's (2014) study found that there was no significant improvement in collocation recognition/recall for frequency exposure between 6 and 12 times, and Pellicer-Sánchez's (2015) study did not identify any significant difference between 4 and 8 times frequency exposure to collocations while reading, which means that frequent exposure alone might not necessarily lead to any improvement in collocation recognition; other factors might play a role in learning.

Cognitive Factors

Those researchers who support the explicit teaching approach focus on instruction, vocabulary learning strategies in particular, as a significant factor that might improve collocation recognition. Mian (1988) talks about the "transfer of training," that is, language errors as a result of incorrect instruction (pp. 33–34). For example, students are taught to memorize words one-by-one, not necessarily related to one another, and they continue to do so when learning collocation patterns, which leads to collocational errors. The authors of textbooks and methodologies on teaching collocations (Gitsaki, 1996; Lackman, 2011; McCarthy & O'Dell, 2005) also draw attention to the importance of learning/perceiving words in combinations when reading and listening.

The factor of learning words in chunks or one-by-one is directly related to recognizing/perceiving word patterns that is usually explored in those studies that examine collocation recognition using reading tasks. Some of these studies argue for drawing the learners' attention to the collocation form and function while reading (Abedi & Mobaraki, 2014; Nesselhauf, 2003; Sonbul, 2015), while others prioritize meaning (Attar & Allami, 2013; Lewis, 2000), and yet other researchers consider both as significant (Sinclair, Jones, & Daley, 2004; Woolard, 2005). Those studies that discuss learning collocations incidentally while reading (Pellicer-Sánchez, 2015; Szudarski, 2012) also note that reading fluency usually correlates with higher levels of collocation competence.

Research Question

Although the existing studies have touched upon some factors that can potentially influence collocation recognition, it still remains unclear how these factors might interact and which could be more significant than others. The

present study addresses this gap by looking into participants' recognition of false collocations, and provides an answer to the following research question:

> What are the differences and similarities of collocation recognition by monolinguals (speakers of English as L1) and bilinguals (speakers of English as L2), and what factors might be responsible for these differences/similarities?

Study Design

This study used a mixed-methods research approach that followed the "exploratory sequential design" (Cresswell & Plano Clark, 2011, p. 69), that is, the results and emerging patterns of the first, small-scale, qualitative stage will form a basis for the second, larger-scale, quantitative stage. The pilot study at the first stage included a collocation-recognition-targeted test, partially modelled after Jafarpour et al. (2013), McCarthy and O'Dell (2005), McKinlay and Hastings (2007), Nguyen and Webb (2016), and Smith (2005), and a posttest interview. The purpose of this stage was to identify the patterns and strategies participants used to select "odd-sounding" word combinations and build a hypothesis about potential factors that might have influenced this recognition. At the second stage, a large-scale quantitative study implemented the same test and posttest survey to identify and compare factors influencing false collocation recognition for the English as L1 and L2 participants. The decision to compare these two populations was motivated by the necessity to understand how/if the mechanism of the ESL learners' recognition of false collocations differs from that of speakers of English as L1 and thus identify factors that might be predictors or mitigators of learning success. Previous studies (Conklin & Schmitt, 2008; Jiang & Nekrasova, 2007; Underwood, Schmitt, & Galpin, 2004) were able to identify some influential factors by comparing two populations of participants, and this research followed in their wake.

Stage 1

Participants

Three female and two male university and college students from different L1 backgrounds in the age range of 21–26 (Table 1) were divided into speakers of English as L1 and speakers of English as L2 (emergent and advanced bilinguals) on the basis of the time they spent in an English-speaking country and age of onset. Emergent bilinguals are defined as individuals who started learning English relatively late and gradually transition to the state of full bilingualism (Menken, 2013), with their primary language still being dominant in all spheres of communication, while advanced bilinguals are those individuals who started learning English earlier, spent more than five years in an

English-speaking country, and whose English proficiency level approaches their primary language and even becomes dominant (Department for Education and Skills, 2006; Office for Standards in Education, Children's Services and Skills, 2005).

Table 1

Stage 1. Participants

Participants	English language proficiency	Age of onset of learning English	Length of residence in the English-speaking environment
Interviewee X	Native speaker	Not applicable	Not applicable
Interviewee Y	Advanced bilingual (Russian/English)	7	One year of studying in the USA; immigrated to Canada at the age of 17; by the time of the interview, had lived in Canada for 10 years
Interviewee Z	Emergent bilingual (German/English)	14	Came to Canada for 6 months as an exchange student; by the time of the interview, had lived in Canada for 4 months
Interviewee U	Emergent bilingual (German/English)	10	Internship in England for 0.5 year; came to Canada for a year as an exchange student; by the time of the interview, had lived in Canada for 5 months
Interviewee W	Emergent bilingual (Chinese/English)	12	Came to Canada at the age of 18; by the time of the interview, had lived in Canada for 2.5 years

Instruments

The recognition-based test was an approximately 500-word-long text (Appendix A), which represented five neutral register paragraphs selected from Canadian news sources such as the Canadian Press, the *Globe & Mail*, and the CBC. The paragraphs were written on a range of everyday life topics from weather forecast to personal narrative and did not have any specialized vocabulary. The test items (Appendix B) were 18 general English collocations consisting of two consecutive notional words, in some cases connected by a functional preposition, which were selected on the basis of their morphosyntactic structure, namely Verb + Noun and Adjective + Noun constructions, which have been reported as the most problematic for learning in several studies (Boers, Demecheleer, Coxhead, & Webb, 2014; Eyckmans, 2009; Gyllstad, 2009; Jaén, 2007; Koya, 2005; Pellicer-Sánchez, 2015). For the purpose of the study, these 18 collocations were deliberately modified. One of the words composing a collocation (either a node or a collocate) was changed to its synonym because, according to Cieślicka (2015), many collocational errors are the result of either direct word-for-word translation or paraphrasing related to word substitution by its synonym. This test format measured recognition

of false collocations and was modelled after Jafarpour et al. (2013), McCarthy and O'Dell (2005), and McKinlay and Hastings (2007). The false counterparts were run through the British National Corpus (BNC) search to ensure their nonfrequency. The test was piloted with three native speakers of English from different academic and nonacademic backgrounds who assessed the text and test items in terms of their "non-nativeness" or "odd-sounding" patterns, neutrality and difficulty of the reading passages, and general understandability of vocabulary.

Semistructured posttest interview questions were split into several parts. The first part addressed the immediate posttest experience and word-selecting and text-reading strategies.

The second part of the interview examined the participants' demographics and language background. The third part focused on their English language proficiency, self-assessment of the four language skills (listening, speaking, reading, and writing), and their vocabulary learning strategies with a focus on word combinations.

Data Collection Procedures

Participants were asked to read the text only once and underline 2-to-3-word combinations that stood out to them as not quite English-like ("odd-sounding"). They were not provided with any prior information about collocations to ensure spontaneity of recognition. The participants were given approximately 8 minutes for test completion (1.5 minutes per one paragraph). The timing corresponded to the average proofreading speed of 200 words per minute as indicated by Ziefle (1998), yet since participants needed to look for specific details, they were given slightly more time. The participants were timed, and although it took emergent bilinguals approximately three minutes longer to read the text, all five participants finished reading in 8 minutes and indicated that they read either "at their normal speed" or "slower." Immediately after taking the test, they were interviewed on the basis of the semistructured questions outlined above, and their answers were audio-recorded.

Data Analysis

When calculating the total test scores for the five respondents, the responses were divided into three categories: false collocations identified by the respondents, those word combinations (including missed false collocations) that participants wanted to select but did not, and the word combinations that were actually correct and yet underlined by the participants as wrong. Immediately after taking the test, participants were asked to explain their choices; their answers ranged from purely intuitive to prior knowledge and/or analysis-based. After that, the data analysis mostly focused on the five interview transcripts. In order to identify patterns and themes emerging in the participants' responses, the descriptive coding linked collocation recog-

OLGA MAKININA

nition to such factors as the primary language background; English proficiency (Groom, 2009; Li & Schmitt, 2010); L1-L2 inter-influence (Irujo, 1986; Liao, 2010; Millar, 2011); vocabulary learning strategies (Lewis, 2000; Mian, 1988; Wray, 2002); the quality and quantity of the language input (Alsakran, 2011; Gitsaki, 1999; Groom, 2009; Wolter & Gyllstad, 2013); the age of onset of learning English (Granena & Long, 2013; Wray, 2002); the length of residence in an English-speaking country (Schmitt, 2000); the predominant language of communication (Wang & Shih, 2011); reading fluency; the focus of attention (criteria of word choice); and attention span while reading (Sonbul, 2015).

Results

The scores on the test were higher for the monolingual speaker of English (83.33%) and the advanced bilingual speaker of Russian/English (72.22%) and lower for the emergent bilinguals, which demonstrated that language proficiency level impacts collocation recognition. However, the scores for the emergent bilinguals did not correlate with the length of their residence in an English-speaking country because Interviewee Z had the shortest stay in Canada (4 months) and the highest score (55.55%), while Interviewee W had the longest stay (2.5 years) and the lowest score (22.22%). The scores also did not correlate with the age of onset of learning English (Interviewee Z, who had the highest score, had started learning English later than Interviewees W and U) or self-assessment of the English language proficiency (Interviewee Z assessed her overall knowledge 2-3 out of 5; while Interviewees U and W gave themselves higher scores).

The research also identified several groups of factors that might have influenced recognition of false collocations.

1. Interlinguistic factors, namely L1-L2 inter-influence, code-switching, and the predominant language for communication.

L1-L2 inter-influence and code-switching. Four bilingual speakers acknowledged that their English was influenced by their primary language; they differed, however, in their reports of the extent of this influence. For one emergent bilingual, his primary language influenced English more often; the advanced bilingual reported that although his mother tongue, Russian, sometimes influences his writing in English ("positioning of certain words or sentences"), it is mostly English that influences his speaking in Russian. There was also a positive relationship between the amount of code-switching and inter-language influence and the length of residence in an English-speaking country. Thus, Interviewee Z almost never code-switched within one conversation ("It really hurts my brain"), while Interviewees U, W, and Y often code-switched from English to their primary language and back in speaking and sometimes in writing (informal emails).

Language of communication. Emergent bilinguals stated their primary language as their main language for communication. However, the extent

to which English plays a role in their thinking and communication differs and correlates with their language proficiency, the length of residence in the English-speaking country, and the amount of interactions with native speakers. Thus, Interviewee Z, who had spent 4 months in Canada and had most of her friends among German-speaking and international students, admitted that she had to make an effort not to mentally translate from German to English. Interviewee U, who spent 0.5 year in England and consciously forced herself to make friends with English-speaking people, stated that she communicated in English 50% of the time, while Interviewee W, who had spent 2.5 years in Canada and had an equal number of English and Chinese-speaking friends, communicated in English 70% of the time. Interviewee Y noticed that his predominant language of thinking and speaking was English, and the role of Russian diminished each year. At the time of the interview, he preferred to speak English even with his Russian-speaking friends.

2. Extralinguistic factors such as the frequency and quality of the input.

The four bilinguals noticed that their fluency and efficiency in English had improved since they had come to an English-speaking country and started communicating with native speakers of English. They also claimed that their classroom instruction and communication in English differed significantly from what they met with in Canada.

> Interviewee Z: *At the airport I [couldn't] ask stewardess for a drink because I didn't know how to say it. It is a completely other situation. Because I just spoke English in my classroom or I watched movies in it, but I never had a real conversation.*

3. Cognitive factors ("reception strategies").

Vocabulary learning strategies. The interviewees stated that at school they learned individual words that were either unrelated or related thematically or grammatically (parts of speech). Interviewees noted that their word learning strategies had changed since they had come to Canada. For example, Interviewee U expressed her preference for "expressions" and words in context.

The focus of attention while reading (criteria of word choice). The participants considered different criteria for selecting a non-English-like word combination. For example, the speaker of English as L1 and the advanced bilingual looked at the "word position in the sentence," the "structure of the word itself," and the overall "sentence flow," while emergent bilinguals mostly focused on the word meaning.

Attention span. Participants had different reading strategies. For example, the participants with higher test scores read for "the fluency of the whole sentence," while emergent bilinguals with lower test scores preferred reading word-by-word.

Discussion

In the context of the pilot study that focused on the qualitative analysis of five interviews with one speaker of English as L1 and four emergent-to-advanced bilinguals, those factors that were addressed in the literature review were explored: interlinguistic (L1-L2 language inter-influence, code-switching, and the predominant language of communication), extra-linguistic (the frequency and quality of the language input), and cognitive (vocabulary learning strategies, criteria of word choice, and attention span). It was also identified that these factors were not quite the same for monolinguals and bilinguals, and for advanced and emergent bilinguals. In terms of cognitive factors ("reception strategies"), such as the focus of attention (criteria of word choice) and attention span while reading, we can see the largest difference. While the advanced speaker of English as L2 and the speaker of English as L1 focused on the overall sentence structure and flow, emergent bilinguals who had lower test scores examined the meanings of individual words. Additionally, the participants with higher test scores read whole sentences, while the participants with lower test scores focused on each word separately. This might be explained by the reading fluency that is achieved through a higher level of mastery of formulaic language (Ellis, Simpson-Vlach, & Maynard, 2008; Wray, 2002). Additionally, those bilinguals who had higher test scores demonstrated more similarities with the speaker of English as L1 in terms of their communicative and cognitive strategies. They indicated that they mainly communicated in English, which corresponds to the findings of Wang and Shih's (2011) study on the relationship between idiomatic competence and English as the predominant language of communication, and were highly motivated and not afraid to step out of their comfort zone when interacting in English. These participants also code-switched more often (Cieślicka, 2015) and preferred to learn words in combinations (Gitsaki, 1996; Lewis, 2000; Wray, 2002). While the pilot study outlined a preliminary map of potentially significant factors, the subsequent quantitative study focused on comparison of these factors on a larger scale between the two populations of speakers of English as L1 and L2.

Stage 2

Participants

Table 2 shows the average distribution of age, gender, and additional (for speakers of English as L1) or primary (for speakers of English as L2) languages across the two groups of university/college undergraduate and graduate students.

	English as L1	English as L2
Total number of participants	50	43
Age	17–29	16–38
Predominant age group	18–19 (40 participants)	18–19 (22 participants)
Gender	36 males, 14 females	30 males, 13 females
Additional/primary languages	31 participants spoke French; 2 participants spoke Arabic; and one participant spoke Chinese	17 participants spoke Chinese; 11 participants spoke Arabic; 4 participants spoke French; and 11 participants spoke other primary languages (i.e., German, Hindi, Portuguese, and Urdu)
Age of onset of learning English	Not applicable	1–32
Length of residence spent in an English-speaking country	Not applicable	4.35 years

As we can see, the predominant age group across the speakers of English as L1 and L2 was 18–19 years old, and almost two thirds of the population was male. The majority of English as L2 speakers started learning English between 5 and 12 years old (27 participants or 62%) and on average had spent less than 5 years in an English-speaking country (34 participants or 79%).

Instruments

Participants were administered the same text as participants in the first stage of the research, and immediately after the test they had to complete a survey (Appendix C). The survey questions targeted those factors that had emerged as important in the previous studies on collocation recognition discussed in the literature review section and had been explored in the pilot study at Stage 1 of this research. The first part of the survey collected information on the participants' demographics and language background; the second part examined the posttest experience and strategies of selecting incorrect (non-English-like-sounding) word combinations; and the third part addressed L1 and L2 inter-influence and vocabulary learning strategies. The survey questions represented closed questions (multiple choice, Likert scale, and "yes/no" questions); however, participants were encouraged to comment on each question.

Data Collection Procedures

Participants (both speakers of English as L1 and L2) were given 8 minutes to read the text once only and underline "odd-sounding," non-English-like 2-to-3-word combinations, and 15 more minutes to complete the survey, as this

test and survey timing had already proved sufficient for the participants of the pilot study at the first stage of the research. They were not provided with any prior information about collocations to ensure spontaneous recognition of false collocations.

Data Analysis

The preliminary analysis divided the responses into three categories: false collocations that were identified by the respondents, missed false collocations, and the actually correct word combinations that were underlined by the participants as wrong. The subsequent steps implemented the SPSS software Version 22. The preliminary descriptive statistics with kurtosis and skewness in the ±2 range indicated normality of the data. As the purpose of the research was to identify those factors that might impact test scores (dependent variable) in the groups of speakers of English as L1 and L2, an independent samples t-test first compared the test scores across the two groups (speakers of English as L1 and L2). Then, the two-way between-groups analysis of variance (ANOVA) identified the relations between the test scores and such independent categorical variables as an additional/primary language, reading speed, criteria of word choice ("form," "meaning," "form and meaning," and "other"), and attention span while reading ("word-by-word." "combinations of 2–3 words," "combinations of 5 and more words," and "sentences"). Next, the one-way ANOVA focused on the factors that were specific for the group of speakers as L2 only, such as their English language proficiency, the predominant language of communication, the L1 influence on English, and vocabulary learning strategies. Finally, the Pearson product-moment correlation coefficient (PPMCC) analysis determined the strength of association between the test scores and such independent continuous variables as the English as L2 participants' language proficiency self-assessment, the age of onset of learning English, and the time they spent in an English-speaking environment.

Results

The significant difference in the scores for speakers of English as L1 (M = 49.77, SD = 20.99) and L2 (M = 27.52, SD = 16.70); t (93) = 5.67, and p = .000 proved that English as L1 has an impact on the collocation recognition. That explains why speakers of English as L1 had higher scores on the collocation recognition test in comparison to speakers of English as L2, and corresponds to the studies by Gitsaki (1999), Groom (2009), Li and Schmitt (2010), and Wray (2002), who stated that at lower proficiency levels, learners could recognize and use collocations less effectively.

For speakers of English as both L1 and L2, the cognitive factor of the focus of attention while reading (criteria of word choice) appeared influential because there was a statistically significant effect of this factor on the test scores [F(3, 85) = 2.85; p = .042), and the effect size was moderate (eta squared = .09). Post hoc comparisons that used the Tukey HSD test showed that the

mean score for the "other" condition of choice (M = 53.53, SD = 26.32) was significantly different from the "word form" condition of choice (M = 30.68, SD = 22.33) because the Sig. value was .008. Those 7 respondents (14%) in the group of speakers of English as L1 and 4 respondents (9%) in the group of speakers of English as L2 who focused on something else than the word form or meaning obtained the highest test scores, and those 11 individuals (22%) in the group of speakers of English as L1 and 10 individuals (23%) in the group of speakers of English as L2 who focused only on the word form had the lowest test scores. However, post hoc comparisons also demonstrated that the "word meaning" condition of choice (M = 35.41, SD = 19.44) and "both word meaning and form" condition of choice (M = 41.84, SD = 20.15) did not significantly differ from the "other" and "word form" conditions; the largest number of participants (28 respondents [56%] and 17 respondents [39.5%] respectively) focused on both word meaning and form when selecting incorrect word combinations. Therefore, the results suggest that the focus of attention on "other" (identified by the respondents as "flow," "internal feeling," and "sentence structure") is an important influential factor that might be related to intuition as discussed in Moon's (1998) and Forsberg and Fant's (2010) research; however, it needs to be further explored due to a small number of participants actually selecting this option. Additionally, since the largest number of participants with higher scores focused on both word form and meaning, this factor might also positively influence test scores, which corresponds to Lewis (2000), and Sinclair et al. (2004), who argue that form and meaning are indivisible parts of the whole.

Two other cognitive factors—attention span while reading and reading speed—did not interact with the test scores in the groups of speakers of English as L1 and L2. Most participants preferred reading combinations of words (23 respondents [46%] in the group of speakers of English as L1 and 18 respondents [41.8%] in the group of speakers of English as L2) to word-by-word reading (4 respondents [8%] in the group of speakers of English as L1 and 3 respondents [6.9%] in the group of speakers of English as L2), which indicates the ESL learners' awareness of the importance of reading and perceiving text in chunks. However, the results of the two-way between-groups ANOVA showed that these differences were not significant, as there was no statistically significant effect of attention span on the test scores [$F(3, 84)$ = 1.16, p = .328], and no interaction effect between the English as L1 factor and the factor of attention span [$F(3, 84)$ = .557, p = .645]. Similarly, the analysis showed that differences in the reading speed (slow, normal, or fast) did not interact with the test scores and were not statistically significant [$F(2,86)$ = 4.02, p = .600]. Predictably, the Pearson product-moment correlation coefficient analysis demonstrated that the factors of age in the group of speakers of English as L1 (r = .077, p = .594) or L2 (r = .041, p = .792) and gender in the group of speakers of English as L1 (r = -.041, p = .778) or L2 (r = .034, p = .830) also did not appear to be affecting the test scores in either of the groups.

OLGA MAKININA

We can see that almost none of the factors that previous studies had identified as important appeared to significantly impact test scores in the group of speakers of English as L1; however, several additional factors played a role in recognition of false collocations by speakers of English as L2. One such factor was, unsurprisingly, English as the predominant language of communication [$F(2, 39) = 5.57$, $p = .007$]. The descriptive statistics results indicated that 17 participants (39.5%) predominantly communicated in their primary language, while 16 participants (37%) used a mix of both languages and only 9 participants (20.9%) predominantly used English. Post hoc comparisons using the Tukey HSD test showed that the largest mean score for "English as a predominant language of communication" ($M = 43.20$, $SD = 14.90$) was significantly different from the lowest mean for "predominantly L1 for communication" ($M = 23.52$, $SD = 15.02$) and the mix of both languages ($M = 24.64$, $SD = 15.70$), and the difference in the mean scores between these three groups was large (eta squared = .220). The results suggest that advanced bilinguals in whom English takes over their primary language could have an advantage over emergent bilinguals in terms of the higher test scores. This finding agrees with Wang and Shih's (2011) research that discussed the importance of English as the predominant language of communication and thinking.

The factor of primary languages was also significant for speakers of English as L2 [$F(3,39) = 5.41$; $p = .003$]. The difference in the mean scores between the groups of speakers of Arabic, Chinese, French, and other languages was large (eta squared = .294), and post hoc comparisons using the Tukey HSD test identified that the mean scores for 17 speakers of Chinese as L1 (39.5%) ($M = 21.23$; $SD = 14.46$) and for 11 speakers of Arabic as L1 (25.5%) ($M = 21.2$; $SD = 13.79$, $p = .013$) were significantly lower than the mean scores for 11 speakers of additional languages, such as German, Hindi, Portuguese, and Urdu (25.5%) ($M = 41.40$; $SD = 15.58$; $p = .006$). This might signify that speakers of Arabic and Chinese as L1 may experience more difficulties when learning to recognize collocations.

Another important finding is the statistical significance of the English as L2 speakers participants' self-assessment using the Likert scale. of their English language proficiency, first in comparison to their primary language proficiency, and then on its own. Those 16 respondents (37%) who indicated their English language proficiency was "as good as their primary language proficiency" ($M = 36.45$, $SD = 17.32$) had higher test scores than those 26 respondents (60%) who stated their English language proficiency was "worse than their primary language proficiency" ($M = 23.07$, $SD = 14.71$); [$t(40) = 2.674$, $p = .011$]. Additionally, the Pearson product-moment correlation coefficient analysis showed that there was a statistically significant relation between how participants evaluated their proficiency level and their test scores ($r = .383$, $p = .010$), which means that the higher their self-assessment was, the higher test scores they had (Figure 1).
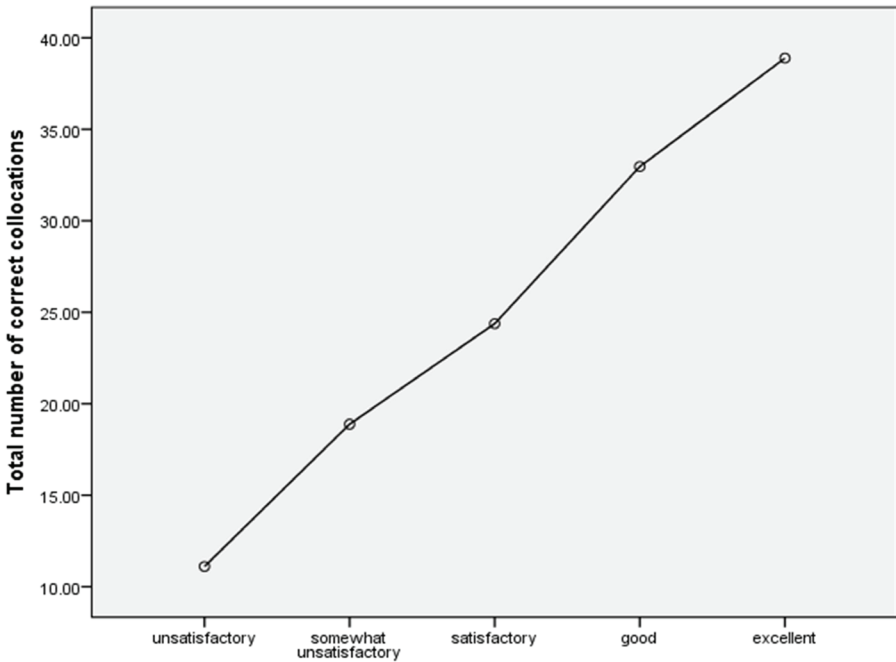
Figure 1: English as L2 Speakers' Self-Assessment of Their English Language Proficiency on the Likert Scale

Unlike the research by Edele, Seuring, Kristen, and Stanat (2015) that demonstrated significant biases in the participants' self-assessment, the finding of this study suggests that ESL learners might be capable of adequate self-assessment of their language proficiency, which corresponds to the level of their collocation competence, and therefore, as Guduru (2014) argues, developing their self-assessment skills might improve their vocabulary learning skills.

Finally, vocabulary learning strategies appeared to be a factor contributing to the recognition of collocations, as there was a statistically significant difference between learning words in sentences/combinations (M = 29.83, SD = 17.23) or one-by-one (M = 19.99, SD = 9.51) [t(29.23) = -2.1, p =.036]. This confirms findings of numerous other classroom-based studies (Gitsaki, 1999; Lewis, 2000; Mian, 1988) that argue about the necessity of learning words in context and chunks.

The age of onset of learning English (r = -.161, p = .321) and the primary language influence on English (r = -.175, p = .286) were not found as contributing to recognition of false collocations. However, according to the Pearson product-moment correlation coefficient analysis, the length of residence in an English-speaking environment was a factor that correlated with the test scores (r = .318; p = .043). This moderate positive correlation suggests that as

OLGA MAKININA

the length of residence in an English-speaking country increases, the total test scores might also potentially increase, which corresponds to the studies by Alali and Schmitt (2012), Durrant and Schmitt (2010), and Wray (2002). However, the moderate correlation might also signify that as Forsberg Lundell and Lindqvist (2014) found out, it is not only the length of residence in an English-speaking country but also other factors such as motivation or aptitude that might play a role in collocation mastery.

Discussion

This quantitative study identified statistically significant factors that impact recognition of false collocations in the groups of speakers of English as L1 and L2, and showed that, although not quite the same, the collocation recognition process follows a similar pattern in case of L1 and L2. This finding confirms the results of the small-scale pilot study at Stage 1 of the research that outlined and described each of these factors in more detail. More specifically, the cognitive factor of the focus of attention on "general flow," "sentence structure," and "internal feeling" when reading and completing the test correlated with higher test scores. These results somewhat oppose the previous research by formalists who argue for the importance of form and grammar (Stefanowitch & Gries, 2003), as well as the supporters of the lexical approach (Lewis, 2000) who stand for the priority of meaning. It seems that in order to recognize collocations, ESL learners might need to step away from morphosyntactic rules and focus on the general fluency and "feeling" of the language, that is, as Forsberg and Fant (2010) and Moon (1998) stated, develop an intuitive and holistic way of perceiving formulaic units. In addition, although some studies (Lewis, 2000; McCarthy & O'Dell, 2005) claimed that reading/listening and comprehending words in chunks might contribute to a more effective collocation recognition, neither of the cognitive processing strategies—such as attention span (reading the text word-by-word, in word combinations, or whole sentences) or reading fluency in general, which was addressed in Sonbul's (2015) study—seemed important in terms of the test scores for the two groups of participants. This might partly be due to the measurement format of the self-report and self-assessment survey that, according to Sitzmann, Ely, Brown, and Bauer (2010) and Edele et al. (2015), might not be effective when measuring cognitive processes. Nevertheless, speakers of English as L1 and speakers of English as L2 with higher test scores did exhibit certain similarities in the way they approached the reading task as most participants preferred to read the text in chunks rather than word-by-word.

Several additional factors emerged as significant in the group of speakers of English as L2. First and foremost, the interlinguistic factor of English as a preferred and predominant language of communication correlated with higher test scores. This finding agrees with Wang and Shih's (2011) research that reinforces the idea of English as the language for thinking and communication increasing formulaic language fluency and proficiency. It also

indicates the importance of the English-speaking environment for learning how to recognize collocations, which was promoted in the studies arguing for language immersion and implicit teaching (Alali & Schmitt, 2012; Durrant & Schmitt, 2010; Pellicer-Sánchez, 2015). This result is further confirmed by another finding of the present study, namely the significance of the extralinguistic factor of years/months that participants spent learning English, and living and working in an English-speaking country. Because most participants had spent less than five years in an English-speaking country, only a moderately positive correlation between this factor and the test scores was identified. However, according to Alsakran (2011), Groom (2009), and Wolter and Gyllstad (2013), the increase of time might potentially improve collocation recognition skills. Another cognitive factor that positively impacted test scores was vocabulary learning strategies. As discussed in the studies by Gitsaki (1996, 1999), Lewis (2000), Mian (1988), and Smith (2005), those individuals who are used to learning words in combinations and patterns might exhibit higher levels of collocation competence.

Some factors that were identified as significant in the previous research, such as the interlinguistic factor of the primary language's influence on English (Liao, 2010; Smith, 2005; Wolter & Gyllstad, 2011) and the extralinguistic factor of the age of onset of learning English (Granena & Long, 2013; Wray, 2002), did not influence the test scores in the group of L2 speakers in the present study. This paradox might be partially explained by the fact that most of the participants started learning English between the ages of 5 and 12, while according to Granena and Long (2013), a younger age might have helped to develop their language intuition and holistic processing and thus positively correlate with recognition of false collocations. Another possible explanation is the participants' prior learning environment. Most of them started learning English in a non-English-speaking country and did not have a total immersion until much later, while such an experience could have positively impacted their collocation skills as discussed in Alsakran's (2011), Groom's (2009), and Wolter and Gyllstad's (2013) research.

Similarly, no statistically significant differences were identified in terms of the participants' age and gender. While it is not surprising that the participants' gender did not impact their ability to recognize incorrect collocations, the age factor is somewhat more important, since some researchers (Han, 2004; Schmitt, 2010; Wray, 2008) discussed the problem of vocabulary fossilization as related to later age of language learning. In contrast to these studies, the present research seems to suggest that collocation recognition might not necessarily be prone to fossilization.

Finally, another important finding identified the participants' ability to accurately assess their English language proficiency since the score they had given themselves on the Likert scale from 1 (*unsatisfactory*) to 5 (*excellent*), as well as their self-assessment of their English language proficiency in comparison to their primary language, positively correlated with their collocation test

scores. This finding needs further exploration, as some studies, for example, Guduru (2014) and Kayler and Weller (2007), argue for using self-assessment in the classroom, but other studies, such as Edele et al.'s (2015) research, warn against relying on the participants' self-assessment as a learning tool because of potential biases.

Conclusion, Limitations, and Future Directions

This three-months-long two-stage mixed-methods research identified statistically significant factors that influence recognition of false collocations by means of a recognition-based test with 18 Verb + Noun and Adjective + Noun collocation items and a posttest interview (at the first research stage) and self-assessment-based survey (at the second research stage). Although the relatively small number of the test items and the self-evaluative component of the survey might have limited the scope of the study to some extent, valuable information on the similarities and differences of the recognition process by speakers of English as L1 and L2 as well as the factors of relevance was obtained. It was determined that two groups of participants (speakers of English as L1 and speakers of English as L2 with higher test scores) shared one common influential cognitive factor, namely the focus of attention on "sentence structure," "flow," "internal feeling," and both word form and meaning, while completing the reading test and searching for the "odd"-sounding word combinations. This signifies that more advanced bilinguals might master some word recognition/processing strategies that are routine skills for speakers of English as L1, and that, in turn, could lead to the improvement of their collocation recognition competence.

Other factors of influence that characterize ESL speakers only had already been explored by other researchers, yet, to the best of my knowledge, were not combined and compared in the context of one comprehensive classification. This classification includes interlinguistic factors, such as English as a predominant language of communication (Wang & Shih, 2011) and the participants' self-assessment of their English language proficiency, the extralinguistic factor of the length of residence in an English-speaking country (Schmitt, 2000), and the cognitive factor of vocabulary-learning strategies (Gitsaki, 1996; Lewis, 2000; Wray, 2002). This finding speaks in favor of the implicit teaching of collocations in the context of language immersion, learning how to perceive words in chunks rather than one-by-one, and using self-reflection and self-assessment as a part of vocabulary studies.

At the same time, some factors that were significant in previous studies, such as interlinguistic factors of L1-L2 inter-influence (Bylund et al., 2012; Irujo, 1986; Liao, 2010; Millar, 2011) and the primary language's influence on English (Liao, 2010; Smith, 2005; Wolter & Gyllstad, 2011), and the extralinguistic factor of the age of onset of learning English (Granena & Long, 2013; Wray, 2002, 2008), were not identified as influential in this study. This might

be due to the fact that most ESL participants in this study belonged to the same age group, started learning English btween ages 5 and 12, and spent less than five years in an English-speaking country. For future studies, it could be useful to compare participants from different ages and prior language learning backgrounds. The cognitive factors of attention span and reading speed were also not identified as significant; however, it is possible that more varied instruments that go beyond the self-assessment survey questions—for example, implementing reaction time tests with a focus on reading errors or pauses as discussed in Rubin (2013) or short-term memory span tests (Jackson & McClelland, 1979)—might have led to different results, which need to be further explored. Another potential limitation is the focus of this study on the external factors rather than internal semantic, morphostructural, and distributional characteristics of collocations as lexical units. In future, it might be useful to conduct a similar study that would focus on collocation-specific factors as impacting its recognition.

Acknowledgments

The Author

Olga Makinina is a PhD candidate in the Applied Linguistics and Discourse Studies program at Carleton University. Her research interests include formulaic language, second language acquisition, and code-switching. She has been teaching English as a second language and Russian in academic institutions in Canada, Eastern Europe, and the USA.

References

Abedi, Z., & Mobaraki, M. (2014). The effect of grammatical collocation instruction on understanding ESP texts for undergraduate computer engineering students. *Journal of Language Teaching and Research, 5*(3), 631–641. https://doi.org/10.4304/jltr.5.3.631-641

Alali, F. A., & Schmitt, N. (2012). Teaching formulaic sequences: The same as or different from teaching single words? *TESOL Journal, 3*(2), 153–180. https://doi.org/10.1002/tesj.13

Alsakran, R. A. (2011). *The productive and receptive knowledge of collocations by advanced Arabic-speaking ESL/EFL learners* (Master's thesis). Colorado State University. Retrieved from ProQuest Databases. (Accession number 1497925)

Attar, E. M., & Allami, H. (2013). The effects of teaching lexical collocations on speaking ability of Iranian EFL learners. *Theory and Practice in Language Studies, 3*(6), 1070–1079. https://doi.org/10.4304/tpls.3.6.1070-1079

Barnbrook, G., Mason, O., & Krishnamurthy, R. (2013). *Collocation: Applications and implications*. New York, NY: Palgrave Macmillan.

Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercise on verb–noun collocations. *Language Teaching Research, 18*(1), 54–74. https://doi.org/10.1177/1362168813505389

Bylund, E., Abrahamsson, N., & Hyltenstam, K. (2012). Does first language maintenance hamper nativelikeness in a second language? A study of ultimate attainment in early bilinguals. *Studies in Second Language Acquisition, 34*(2), 215–241. https://doi.org/10.1017/S0272263112000034

CBC News (June 2, 2013). CP says no risk to public after several cars topple into Wahnapitae River. CBC News. Retrieved from http://www.cbc.ca/news/canada/sudbury/train-derails in-bridge-collapse-near-sudbury-ont-1.1301044

CBC News (September 27, 2013). At least 35 homeless after fast-moving rowhouse fire. *CBC News*. Retrieved from http://www.cbc.ca/news/canada/ottawa/at-least-35-homeless-after fast-moving-rowhouse-fire-1.1870403

CBC News (October 10, 2013). Southern Manitoba to get heavy rain, thunderstorms. *CBC News*. Retrieved from http://ca.news.yahoo.com/southern-manitoba-heavy-rain-thunderstorms 171738166.html

Cieślicka, A. B. (2015). Idiom acquisition and processing by second/foreign language learners. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual figurative language processing* (pp. 208–244). Cambridge, UK: Cambridge University Press.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72–89. https://doi.org/10.1093/applin/amm022

Creswell, J. V., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage.

Department for Educations and Skills [DfES]. (2006). *Excellence and enjoyment: Learning and teaching for bilingual children in the primary years* (Ref. number 0013-2006DCL-EN). Colegate, Norwich, UK: DfES Publications.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research, 26*(2), 163–188. https://doi.org/10.1177/0267658309349431

Edele, A., Seuring, J., Kristen, C., & Stanat, P. (2015). Why bother with testing? The validity of immigrants' self-assessed language proficiency. *Social Science Research, 52*, 99–123. https://doi.org/10.1016/j.ssresearch.2014.12.017

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly, 42*(3), 375–396. https://doi.org/10.1002/j.1545-7249.2008.tb00137.x

Eyckmans, J. (2009). Toward an assessment of learners' receptive and productive syntagmatic knowledge. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 139–152). New York, NY: Palgrave Macmillan.

Firth, J. R. (1957). Modes of meaning. In J. R. Firth (Ed.), *Papers in linguistics 1934–1951* (pp. 190–215). Oxford, UK: Oxford University Press.

Forsberg, F., & Fant, L. (2010). Idiomatically speaking: Effects of task variation on formulaic language in highly proficient users of L2 French and Spanish. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 47–70). New York, NY: Continuum International.

Forsberg Lundell, F., & Lindqvist, C. (2014). Lexical aspects of very advanced L2 French. *Canadian Modern Language Review, 70*(1), 28–49. https://doi.org/10.3138/cmlr.1598

Gitsaki, C. (1996). *The development of ESL collocational knowledge* (Doctoral thesis). Centre for Language and Research, University of Queensland. Retrieved from University of Queensland (UQ) Theses (RHD) at http://espace.library.uq.edu.au/view/UQ:205374

Gitsaki, C. (1999). *Second language lexical acquisition: A study of the development of collocational knowledge*. Bethesda, MD: International Scholars Publications.

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research, 29*(3), 311–343. https://doi.org/10.1177/0267658312461497

Graveland, B. (October 25, 2013).Three children in critical condition, several others hurt, in bus crash near Calgary. *Canadian Press*. Retrieved from http://kitchener.ctvnews.ca/three-children-in-critical-condition-several-others-hurt-in bus-crash-near-calgary-1.1513301

Groom, N. (2009). Effects of second language immersion on second language collocational development. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 21–33). New York, NY: Palgrave Macmillan.

Guduru, R. (2014). An overview of Saudi EFL learners' self-assessment of vocabulary learning strategies. *Language in India, 14*(5), 242–260.

Gyllstad, H. (2009). Designing and evaluating tests of receptive collocation knowledge: COLLEX and COLLMATCH. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 153–170). New York, NY: Palgrave Macmillan.

Han, Z.-H. (2004). *Fossilization in adult second language acquisition*. Tonawanda, NY: Multiligual Matters.

Henriksen, B. (2013). Research on L2 learners' collocational competence and development: A progress report. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis*. Eurosla Monograph Series, 2 (pp. 29–56). Retrieved from http://www.eurosla.org/monographs/EM02/EM02home.php

Henriksen, B., & Stoehr, L. S. (2009). Commentary on Part IV: Processes in the development of L2 collocational knowledge – A challenge for language learners, researchers and teachers. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 224–231). New York, NY: Palgrave Macmillan.

Hughes, G. (April 19, 2012). 150 arrested during escalating Quebec student protest. *Globe and Mail*. Retrieved from https://beta.theglobeandmail.com/news/national/150-arrested during-escalating-quebec-student protest/article4101266/?ref=http://www.theglobeandmail.com&

Irujo, S. (1986). Don't put your leg in your mouth: Transfer in the acquisition of idioms in a second language. *TESOL Quarterly*, *20*(2), 287–304. https://doi.org/10.2307/3586545

Jackson, M. D., & McClelland, L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology, 108*(2), 151–181. https://doi.org/10.1037/0096-3445.108.2.151

Jaén, M. M. (2007). A corpus-driven design of a test for assessing the ESL collocational competence of university students. *International Journal of English Studies, 7*(2), 127–147.

Jafarpour, A. A., Hashemian, M., & Alipour, S. (2013). A corpus-based approach toward teaching collocation of synonyms. *Theory and Practice in Language Studies, 3*(1), 51–60. http://doi.org/10.4304/tpls.3.1.51-60

Jiang, N. A. N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *Modern Language Journal, 91*(3)*,* 433–445. http://doi.org/10.1111/j.1540-4781.2007.00589.x

Kayler, M., & Weller, K. (2007). Pedagogy, self-assessment, and online discussion groups. *Journal of Educational Technology and Society, 10*(1), 136–147.

Koya, T. (2005). *The acquisition of basic collocations by Japanese learners of English* (Doctoral dissertation). Waseda University, Tokyo, Japan. Retrieved from http://www.cs.waikato.ac.nz/~shaoqun/tmp/Honbun-4160.pdf

Krishnamurthy, R. (2005). Collocations. In K. Brown (Ed.), *Encyclopedia of language and linguistics*. Elsevier Science. ISBN 9780080448541

Lackman, K. (2011). *Teaching collocations: Activities for vocabulary building.* Toronto, ON: Ken Lackman & Associates, Educational Consultants. Retrieved from http://www.kenlackman.com/files/Collocations11sample.pdf

Levitzky-Aviad, T., & Laufer, B. (2013). Lexical properties in the writing of L2 learners over eight years of study: Single words and collocations. In C. C. Bardel, B. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis.* EUROSLA Monographs Series, 2 (pp. 127–148).

Lewis, M. (2000). Language in the lexical approach. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 126–154). Boston, MA: Thomson & Heinle.

Li, J., & Schmitt, N. (2010). The development of collocation use academic text by advanced L2 learners: A multiple case study approach. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 23–46). New York, NY: Continuum International.

Liao, E.-H. (2010). *An investigation of cross-linguistic transfer in EFL learners' phraseology* (Master's thesis). Retrieved from ProQuest LLC. (Accession number 3401772)

MacWhinney, B. (2006). Emergent fossilization. In Z. Han & T. Odlin (eds.), *Studies of fossilization in second language acquisition* (pp. 134–156). Clevedon, UK: Multilingual Matters.

McCarthy, M., & O'Dell, F. (2005). *English collocations in use: Advanced*. Cambridge, UK: Cambridge University Press.

McKinlay, S., & Hastings, B. (2007). *Success: Intermediate student's book.* Harlow, UK: Pearson Education.

Menken, K. (2013). Emergent bilingual students in secondary school: Along the academic language and literacy continuum. *Language Teaching, 46*(4), 438–476. https://doi.org/10.1017/S0261444813000281

Mian, A. H. (1988). *A microanalysis of collocation in the interlanguage of Pakistani adults learning English as a second language* (Master's thesis). University of British Columbia, Vancouver, BC. https://doi.org/10.14288/1.0078311

Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*, *32*(2), 129–148. https://doi.org/10.1093/applin/amq035

Moon, R. (1998). Frequencies and forms of phrasal lexemes in English. In A. P. Cowie (Ed.), *Phraseology. Theory, analysis, and applications* (pp. 79–100). Oxford, UK: Oxford University Press.

Nakuma, C. (1998). A new theoretical account of 'fossilization': Implications for L2 attrition research. *IRAL–International Review of Applied Linguistics in Language Teaching, 36*(3), 247–256.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*(2), 223–242. https://doi.org/10.1093/applin/24.2.223

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam, Netherlands: John Benjamins.

Nguyen, T. M. H., & Webb, S. (2016). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research, 21*(3), 298–320. https://doi.org/10.1177/1362168816639619

Office for Standards in Education, Children's Services and Skills [Ofsted]. (2005, July 13). Could they do even better? *Office for Standards in Education, Children's Services and Skills Departmental Report 2008-09*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/238633/7597.pdf

Pellicer-Sánchez, A. (2015). Learning L2 collocations incidentally from reading. *Language Teaching Research, 21*(3), 381–402. https://doi.org/10.1177/1362168815618428

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research, 20*(1), 113–138. https://doi.org/10.1177/1362168814568131

Rubin, G. S. (2013). Measuring reading performance. *Vision Research, 90*, 43–51. https://doi.org/10.1016/j.visres.2013.02.015

Schmitt, N. (2000). *Vocabulary in language teaching.* Cambridge, UK: Cambridge University Press.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. New York, NY: Palgrave McMillan.

Sinclair, J. M., Jones, S., & Daley, R. (2004). *English collocation studies: The OSTI report*. London: Continuum.

Sitzmann, T., Ely, K., Brown, K. G., & Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning and Education, 9*(2), 169–191. https://doi.org/10.5465/AMLE.2010.51428542

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multistudy perspective. *Canadian Modern Language Review*, *64*(3), 429–458. https://doi.org/10.3138/cmlr.64.3.429

Siyanova-Chanturia, A., & Spina, S. (2015). Investigation of native speaker and second language learner intuition of collocation frequency. *Language Learning: A Journal of Research in Language Studies, 65*(3), 533–562. https://doi.org/10.1111/lang.12125

Smith, C. (2005). *The lexical approach: Collocation in high school English language learners* (Doctoral dissertation). George Fox University. (Accession number 3195526)

Stefanowitsch, A., & Gries, S. (2003). Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics, 8*(2), 209–243. https://doi.org/10.1075/ijcl.8.2.03ste

Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing. *Bilingualism: Language and Cognition, 18*(3), 419–437. https://doi.org/10.1017/S1366728914000674

Szudarski, P. (2012). Effects of meaning-and form-focused instruction on the acquisition of verb-noun collocations in L2 English. *Journal of Second Language Teaching & Research, 1*(2), 3–37.

Szudarski, P., & Carter, R. (2014). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics, 26*(2), 245–265. https://doi.org/10.1111/ijal.12092

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 153–172). Philadelphia, PA: John Benjamins North America.

Wang, H. C., & Shih, S. C. (2011). The role of language for thinking and task selection in EFL learners' oral collocational production. *Foreign Language Annals, 44*(2), 399–416. https://doi.org/10.1111/j.1944-9720.2011.01135.x

Webb, S., & Kagimoto, E. (2011). Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics, 32*(3), 259–276. https://doi.org/10.1093/applin/amq051

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics, 32*(4), 430–449. https://doi.org/10.1093/applin/amr011

Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition, 35*(3), 451–482. https://doi.org/10.1017/S0272263113000107

Woolard, G. (2005). *Key words for fluency: Intermediate collocation practice*. Boston, MA: Thomson Heinle.

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

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford, UK: Oxford University Press.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language and Communication, 20*(1), 1–28. https://doi.org/10.1016/S0271-5309(99)00015-4

Ying, Y., & O'Neill, M. (2009). Collocation learning through an 'AWARE' approach: Learner perspectives and learning process. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 181–193). New York, NY: Palgrave Macmillan.

Ziefle, M. (1998). Effects of display resolution on visual performance. *Human Factors, 40*(4), 554–568. https://doi.org/10.1518/001872098779649355

Zyzik, E. (2009). Teaching and learning idioms: The big picture. *CLEAR News, 13*(2). Retrieved from http://clear.msu.edu/clear/newsletter/files/fall2009.pdf

Appendix A. Test

Instructions: Please read the text below and underline the <u>2-to-3-words long combinations</u> that seem "odd," not English-like, or "unnatural-sounding" to you. Please <u>do not</u> underline one word only or a whole sentence. You will be timed. If you finished reading before the assigned time, please <u>do not</u> reread the text again. <u>Read the text only once</u> at your normal speed.

***

Three children were in critical condition in hospital and six others were in serious but <u>balanced condition</u> after the school bus crashed into a truck north of Calgary, Alberta Health Services said. The children <u>fluctuated in age</u> from 5 to 13. "My heart goes out to the students and parents directly impacted by this incident. We are very concerned about the <u>good-being</u> of our school communities," said Susan William, acting superintendent of the Rocky View School Division. "We're not quite sure what the <u>supplying factor</u> to this collision was, whether that's <u>finite visibility</u> or road conditions or what's actually there," the police officers said.
(Graveland, 2013)

***

At least 35 people in several units of a Hurdman-area rowhouse are homeless after a fast moving fire that broke out Friday morning. Eight rowhouse units at 211 Lees Ave. were damaged at a <u>guessed cost</u> of $500,000. When firefighters arrived, the fire had spread to three of the 14 units, and several vehicles parked near the building had also <u>captured fire</u>. One firefighter was taken to the hospital for an <u>unimportant injury</u>, and no one else was hurt. The cause of the fire is not yet known.
(CBC News, September 27, 2013)

***

Environment Canada says <u>thick rain</u> and possibly thunderstorms are on the way Friday and Saturday for western areas of southern Manitoba. The agency has issued a special weather statement for: A storm system over the southwestern US will hit southern Manitoba Friday night. Rain is expected over the southwest part of the province Friday morning and continuing into Friday night, with rainfall amounts of 40 to 60 mm in the southwest Manitoba and <u>taller amounts</u> possible locally as well. <u>Acute thunderstorms</u> are also possible across southern Manitoba Friday afternoon.
(CBC News, October 10, 2013)

***

Canadian Pacific says there is no indication that a train derailment east of Sudbury on Sunday <u>produces a danger</u> to the public or the environment. Images of the scene showed the bridge collapsed and a number of cars carrying containers fell into the Wahnapitae River. Canadian Pacific spokesperson Ed Greenberg says <u>introductory inspection</u> of the containers indicates there are no materials or products that are dangerous. He says Canadian Pacific crews are working with <u>native officials</u> to determine the condition of the containers. The <u>beginning investigation</u> has found one of the rail cars had an unexpected wheel bearing failure that caused the derailment just before the bridge, Greenberg said in an email.
(CBC News, June 2, 2013)

***

Quebec students have been in the streets this spring protesting against tuition increase, which would <u>promote fees</u> 75 percent more over five years but still leave the province with rates lower than the Canadian average. While many students have returned to their classrooms, the <u>proceeding protests</u> have disrupted subway service in Montreal. There were warnings Thursday that the conflict is getting costly—aside from the potential price tag of <u>enlarging the semester.</u> Provincial police said the protests have already cost them $1.5 million just in overtime.
(Hughes, 2012)

Appendix B. List of the Test Items and Their False Counterparts

Test items	False counterparts
Stable condition	Balanced condition
Ranged in age	Fluctuated in age
Well-being	Good-being
Contributing factor	Supplying factor
Limited visibility	Finite visibility
Estimated cost	Guessed cost
Catch fire	Capture fire
Minor injury	Unimportant injury
Heavy rain	Powerful rain
High amount	Tall amount
Severe thunderstorm	Acute thunderstorm
Pose a danger	Produce a danger
Preliminary inspection	Introductory inspection
Local official	Native official
Initial investigation	Beginning investigation
Raise fees	Promote fees
Ongoing protest	Proceeding protest
Extend the semester	Enlarge the semester

Appendix C. Posttest Survey

Part 1

1. Your age: _____
2. Your gender (please circle):   Male   Female
3. Is English your primary language? (please circle)   Yes   No

If you answered "yes" in Question 3, please proceed to <u>Questions 4–7</u> and skip Questions 8–12.
If you answered "no" in Question 3, please proceed to <u>Questions 8–12</u> and skip Questions 4–7.

4. Please name another language except English (if any) that you know best: _____
5. Your knowledge of the language named in **Question 4** is:
   As good as your knowledge of English
   Worse than your knowledge of English
6. How would you assess your knowledge of the language named in **Question 4** on a scale from 1 (unsatisfactory) to 5 (excellent)?
   　　1　　　　　　　2　　　　　　　3　　　4　　　5
   (Unsatisfactory) (Somewhat unsatisfactory) (Satisfactory) (Good) (Excellent)
7. At what age did you begin learning language named in **Question 4**? _____
8. Please name your primary language: _____
9. Your knowledge of English is:
   As good as your knowledge of the language named in **Question 8**
   Worse than your knowledge of the language named in **Question 8**
10. How would you assess your knowledge of English on a scale from 1 (*unsatisfactory*) to 5 (*excellent*)?
    　　1　　　　　　　2　　　　　　　3　　　4　　　5
    (Unsatisfactory) (Somewhat unsatisfactory) (Satisfactory) (Good) (Excellent)
11. At what age did you begin learning English? _____
12. How many years/months did you study English in an English-speaking country (for example, Canada, Great Britain, USA)? _____

Part 2
1. How quickly did you read?
   Slower than usual
   At my normal speed
   Faster than usual
2. What did you focus your attention on when you were choosing the incorrect word combinations?
   Word meaning
   Word form (ex., part of speech, grammar)
   Both
   Other (please explain): _____
3. When you were reading, you mostly focused on:
   Word-for-word individually
   Combinations of two-three words
   Combinations of five or more words
   Sentences

Part 3
1. What language do you mostly communicate in?
   English
   Other (please name): _____
   Both

If in Part 1 you indicated English as your second language, please answer <u>Questions 2 and 3</u>, and skip Question 4.
If in Part 1 you indicated English as your primary language, please answer <u>Question 4 and skip Questions 2 and 3</u>.
2. To what extent does your primary language influence your English? In a few words explain how.
   1   2   3   4   5
   (Never) (Very rarely) (Sometimes) (Often) (Very often)
3. How do you prefer to learn new words in English:
   One-by-one, not necessarily related to one another
   In sentences/combinations with other words
   Other (please state) _____
4. To what extent does another language that you know/learn influence your English? In a few words explain how.
   1   2   3   4   5
   (Never) (Very rarely) (Sometimes) (Often) (Very often)