Chabahar Maritime University
Iranian Journal of English for Academic Purposes ISSN: 2476-3187
IJEAP, (2016) vol. 5 issue. 1 (Previously Published under the title: Maritime English Journal)

Collocational Processing in Two Languages:

A psycholinguistic comparison of monolinguals and bilinguals

[1] **Parastoo Harati Mokhtari**

[2] **Reza Ghafar Samar***

[3] **Gholam Reza Kiany**

Received: 01/04/2016 Accepted: 28/05/2016 Available online: 01/06/2016

Abstract

With the renewed interest in the field of second language learning for the knowledge of collocating words, research findings in favour of holistic processing of formulaic language could support the idea that these language units facilitate efficient language processing. This study investigated the difference between processing of a first language (L1) and a second language (L2) of congruent collocations. A read-aloud task was administered to monolinguals (Persian) and advanced bilinguals (Persian-English) to assess their reaction time (RT) to collocations in comparison to non-collocations. With respect to the frequency-based approaches to the study of collocations, a list of high frequency items and a second list of non-collocations were compiled. It was hypothesised that when bilinguals are at advanced levels of proficiency their processing rate should not differ significantly from monolinguals. The study also examined the assumptions of dual-route system model (Wray, 2002). The data was analysed using linear mixed-effects modelling. The results suggested no advantage for processing of collocations over non-collocations, whereas collocational processing displayed higher response rate for monolinguals in comparison to bilinguals which contradicted our hypotheses and is in line with dual-route model assumptions.

Keywords: collocations; formulaic language; holistic processing; read-aloud; linear mixed-effects modelling (LMM); psycholinguistics

1. Introduction

Knowing a language involves, not only the knowledge of individual words, but also the manner in which they fit together in a form which is mostly refer to as formulaic

[1] Department of Applied Linguistics, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran
Email: p.harati@yahoo.co.uk

[2] Corresponding Author- Department of Applied Linguistics, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran Email: rgsamar@modares.ac.ir

[3] Department of Applied Linguistics, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran Email: kiany_gh@modares.ac.ir

language (Wray, 2002). Acquisition of memorised sequences of language is part of achieving a fluent comprehension and production, in both one's first (L1) and second (L2) languages. In addition, learning to select customarily, from a great many possibilities of words that could be selected, as done by the native speakers (NSs), is an extremely difficult task for even the most proficient non-native speakers (NNSs) (Pawley & Syder, 1983). Regardless of what we are concerned with, be it language learning or language use, multiword units (e.g. idioms, formulaic sequences, and collocations) have shown to be of great importance (Yamashita & Jiang, 2010).

Formulaic language has been considered to be divided into various types, for instance, idioms, collocations, phrasal verbs, lexical bundles, etc. Collocations are claimed to be one of the four most significant connections in the mental lexicon (Aitchison, 2012). Despite their problematic nature, collocations play an indispensable role in language learning and teaching. Erman and Warren (2000), stated that multiword units (including collocations) constitute 58.6% of spoken and 52.3% of the written analysed text. Collocations are considered to be important mostly because they provide information on semantics. When a speaker encounters an ambiguous word, what helps them make the right choice of meaning are the words that collocate with it (Durrant, 2008).

Recently, extensive research has emerged, focusing on the storage and processing of these linguistic units, with specific attention on whether such processing is of holistic or analytic nature. Second and foreign language learning studies also recently suggest that more focus needs to be assigned to formulaic language (Millar, 2011). Empirical research in last few years has mostly confirmed that processing of multiword units is quicker and more accurate, and this is said to be due to the holistic processing of these sequences. These researches, however, have mostly focused on idioms, lexical bundles, and lexical phrases (e.g., Conklin & Schmitt, 2007, 2012; Schmitt, 2004; Weinert, 1995; Wray, 2002; Gibbs, Nayak, & Cutting, 1989; Niemi, & Laine, 2002; Siyanova-Chanturia, Conklin, & Schmitt, 2011), and little attention seems to have been directed toward another important type of multiword units, the collocations (see Yamashita & Jiang, 2010; Wolter, 2014; Wolter & Gyllstad, 2011, 2013; Siyanova & Schmitt, 2008; Gyllstad & Wolter, 2015).

Online tasks have a long history in studying processing of idioms in L1. Some studies have suggested that idioms go through holistic processing and a halt of literal processing (e.g Cacciari & Tabossi, 1988; Cutting & Bock, 1997). Processing advantages have also been observed in some other studies on idioms (e.g Siyanova-Chanturia, Conklin, & Schmitt, 2011), binominals (e.g., Siyanova-Chanturia, Conklin, & van Heuven, 2011), formulaic sequences (e.g Conklin & Schmitt, 2008), and lexical bundles (e.g. Tremblay, Derwing, Libben, & Westbury, 2011) in native and proficient non-native speakers.

Not only is the online processing of collocations new to studies on processing of formulaic language, but it also suffers from receiving little attention in the realm of L2 processing (see Gyllstad & Wolter, 2015; Siyanova & Schmitt, 2008; Wolter & Gyllstad, 2011, 2013; Wolter & Yamashita, 2014; Yamashita & Jiang, 2010). Most of the studies to date, have focused on European languages and have neglected languages that are different in orthography, word-order and writing systems.

Furthermore, all studies on processing of second language formulaic sequences, except one (Sugiura, Fujumura, Yamashita, & Leung, 2013), have examined collocations, drawing comparisons between native speakers of English and L2 speakers of other languages in their second language. So far only an unpublished study by Sugiura et al. (2013), has examined Japanese native speakers (a language different from English in terms of orthography, language family, writing system, etc.) in comparison to Japanese-French second language learners in their second language.

Bearing this in mind, this research brings into question whether processing differences are visible in collocations occurring frequently in the corpora in comparison to word combinations that do not co-occur in the corpora, which are henceforth referred to as non-collocations. Furthermore, the difference in the processing of collocations by monolinguals in comparison to bilinguals are highlighted, taking into account the nature of their first language (Persian). Languages with word-order, orthography, and writing systems like that of Persian, have been neglected in L1 and L2 studies. In addition, the lack of studies comparing monolingual groups with L1s other than English calls for further investigations.

It was hypothesised that collocations would display a processing advantage, which would serve as an indication that these types of word combinations are processed holistically and as a single unit. Also, the processing advantage was theorised to be viewed in both language groups. This was so because the bilingual group were chosen among unbalanced bilinguals who were chosen through a proficiency test and were considered highly proficient as they were all graduate students of English related fields.

In addition, the study aimed to explore the assumptions made by the dual-route (dual-systems) model (Wray, 2002). This model proposes holistic and analytic processing types, which are different at each stage of a person's life. The theory suggests that the processing of L1 may be both holistic and analytic at later stages of life, as it develops during childhood and it is finally balanced in adults, resulting in processing advantage for formulaic sequences in L1. However, L2 learners, at the beginning stages of learning, have not developed the dual systems and hence the holistic and analytic processing is not balanced for processing of formulaic sequences in L2 learners. This is believed to be the reason for L2 learners not showing any processing advantage for formulaic sequences, specifically at earlier stages of learning. So L2 learners are more likely to process language analytically, and hence they are more prone to negative transfer of L1 processing patterns to their L2.

The current study aims to investigate processing of collocations as opposed to non-collocations by L1 and L2 speakers. A large number of the studies have focused on L1 (see Arnon & Cohen Priva, 2013; Arnon & Snider, 2010; Ellis & Simpson-Vlach, 2009; Janssen & Barber, 2012; Shaoul, 2012; Sosa & MacFarlane, 2002; Tremblay, Derwing, Libben, & Westbury, 2011b), and insufficient attention has been devoted to the area of bilingualism and L2 processing, especially when it comes to considering a range of different languages other than English. This study is therefore an attempt at examining the bilingual processing of collocations through employing online techniques. Also, this research aims to identify if the learners of a language, process collocations differently in their non-native language. This is carried out using read-aloud experimental tasks, which have previously been employed for exploring processing of formulaic language types other than collocations.

Also, we intended to consider collocation processing in speakers of Persian for the first time. This may provide a clearer picture of whether or not some of the previous studies done on collocation processing on participants of European languages benefited from having the same language background and also learners having more exposure to the language input. In a study by Wolter (2014), the difference between the English Native speakers and Japanese learners of English was suggested to be explained based on the language having much less in common with English (in terms of orthography and the amount of input that is available to the Japanese learners of English). This was inconsistent with previous research carried out by Wolter & Gyllstad, (2011, 2013) on Swedish language, which came from the same orthographical background and more exposure to English input.

Moreover, due to the lack of research studies comparing the collocational processing of native speakers of a L1 other than English, this research contributes to further the understanding and exploration of collocations in a different L1 context.

2. Method

2.1 Participants

A total of 44 participants took part in this study. A group of nineteen Persian monolinguals with the age range of 19-39 (mean = 24.9) were recruited, as our control group, from among undergraduate and master's students/graduates and employees of Tarbiat Modares, Tehran, and Allameh Tabatabaei Universities, who were studying or had studied non-English fields of study. This group was selected through self-report as monolinguals, where they declared being speakers of Persian only and not knowing other languages. Their level of exposure to language was merely limited to school or general English courses in university, which involved very limited knowledge of grammar, vocabulary, and reading. Also, a second group of twenty-five Persian-English bilinguals with the age range of 22-39 (mean = 26.7) were recruited from MA/PhD students or graduates of Applied Linguistics, TEFL, English Literature, or English Translation of Tarbiat Modares University, Tehran University, or Allameh Tabatabaei University. To make sure that the group fit proficiency requirements of the study (only advanced students), Oxford Placement Test (OPT) was administered to them prior to the experiment. Moreover, both groups reported normal or corrected-to-normal vision and had no problem reading from a computer screen. Only one of the participants was left-handed.

2.2 Item Development

Item development was a pivotal part of this research. To develop the experimental items of the study, we first had to determine what was meant by collocations. The amount of interest in formulaic language in general, and collocations in specific, has resulted in a confusion on how these sequences should be defined, and researchers are far from reaching a consensus on the terminology regarding formulaic language yet.

In defining collocations two major approaches have been adopted historically: frequency-based approach and phraseological approach. The former approaches collocations as words which co-occur frequently. It involves corpus linguistics and application of statistical measures as the frequency of co-occurrence is normally obtained using a corpus or several corpora. With the development of computer-related technologies, this approach has received a lot of attention in second language acquisition studies and corpus linguistics, when working with any type of formulaic sequences (Barfield & Gyllstad, 2009). Within the latter approach, collocations are defined as word combinations in which at least one word in the combination is not semantically transparent. Collocations in this view are classified based on a continuum, from fully transparent to fully opaque (e.g., Gyllstad, 2007).

Each of the approaches has its own benefits and limitations. In this research we defined our version of collocations along the lines of the frequency-based approach and controlled and checked for frequency of collocational items in this research against the corpora. Nevertheless, we did not exclusively work with frequency-based approach as collocational items were also checked for their transparency.

After defining collocations, we had to choose the type of collocations we were going to bring under investigation. In this study, we worked on adjective-noun collocations because this type of collocations was the most amenable when working with Persian and English language and it was the easiest to find counterparts for in both languages.

To compile a list of experimental collocational items, data was obtained from the 100-million-word BNC corpus, which was accessed through the BNCweb (Lehmann, Hoffmann, & Schneider, 2002) corpus tool. First, a list of high-frequency nouns was extracted and a second list consisting of each noun's adjective collocates was also created. From the adjective-noun combinations which had a frequency above 10 per-million-words (normalised or relative frequency count). The collocations were then checked for their mutual information (MI) score (3≤) and only those meeting the requirement were included in the English collocational item list. These collocations were then translated into Persian which were again checked by two linguists, who were also native speakers of Persian and proficient English speakers to make sure there were not significant variations in the translations and collocations were congruent (i.e., having an equivalent first language construction). Only collocations which had an equivalent congruent counterpart in Persian remained in the English collocational item list. Persian collocations were then checked for their frequency in Peykare Corpus (Bijankhan, Sheykhzadegan, Bahrani, & Ghayoomi, 2011). Likewise, their MI score was also calculated. The final list included merely the collocations which passed the criteria and had a counterpart in the other language. These procedures left us with 8 English and 8 Persian collocations in our final list. It is worth noting that we did not include any collocational homophones or (false) cognates. In other words, there were no items for which both the adjective and the noun were homophone or cognates in Persian and English. Also words in both languages were checked in a dictionary to avoid inclusion of an adjective-noun compound word instead of a collocation.

In addition to the collocational items, a list of non-collocational items containing adjectives-noun combinations with low MI scores (i.e., less than 3) were taken as non-collocational baseline items to serve as control items. In order to make the experimental items more comparable, item length (i.e., number of letters) were also controlled for the whole word combination (see Appendix A for the list of items).

2.3 Procedure

A read-aloud experimental task was used to gather RT data for all items in this research. In this type of productive psycholinguistic task, the participants are asked to read the stimuli appearing on the screen as quickly and as accurately as possible. A total of 24 items, including 8 adjective-noun collocations and 8 non-collocations were used as stimuli.

The experiments were presented using DMDX display software (version 5.1.3.4; Forster & Forster, 2003), the read-aloud task was administered to each participant who was asked to read each item aloud as quickly and accurately as possible into a microphone. The read-aloud task included the following steps: (1) Before starting the task, participant's voice was collaborated using "test vox" in the DMDX menu to make sure the sensitivity was good enough to record the response as a trigger in time; (2) First couple of screens on the read-aloud task provided instructions on the read-aloud task, which participant had to press spacebar to move to view the next part of the instruction (position: centred; English font: Courier New, size:12; Persian font: FreeFarsi, size: 14); (3) After viewing all the instructions, participant was invited to press spacebar to start a practice session (position:

centred; English font: Courier New, size:12; Persian font: FreeFarsi, size: 14); (4) The participant was asked to read aloud each phrase as soon as s/he recognised it, and as quickly and accurately as possible (position: centred; English font: Courier New, size:12; Persian font: FreeFarsi, size: 14); (5) A practice session with five practice items (not part of the main test items) followed; (6) The participant was required to press spacebar to start the practice session; (7) A short presentation of a fixation appeared on the screen and the window displaying the stimulus followed (position: centred; English font: Courier New, size:12; Persian font: FreeFarsi, size: 14); (8) As soon as the participant pronounced the stimulus, a new window displaying a fixation appeared briefly, which was immediately followed by a window displaying the next stimulus; (9) The practice session ended after presentation of five item sentences. Any questions after the practice were answered, and the participant began the test session by pressing the space bar. No break was allowed in the middle of the test session; (10) Once the participant finished articulating all the items in the task, a message indicating the end of the experimental task was shown. The participant was asked to press ESC to end the program.3.3.1 Sample Size, Power, and Precision

Along with the description of subjects, give the mtended size of the sample and number of individuals meant to be in each condition if separate conditions were used. State whether the achieved sample differed in known ways from the target population. Conclusions and interpretations should not go beyond what the sample would warrant.

3. Results and Analysis

The initial dataset was cleaned before the analysis with the aim of identifying and removing missing data and outliers to avoid any effects they may have had on the distribution of the data. The recorded data was then analysed using linear mixed-effects modelling (LMM; see Appendix B for a brief inroduction). In order to analyse the data, we used R version 3.2.4 (R Development Core Team, 2014), which is free to download from http://www.r-project.org, and the R packages "lme4" (Bates & Maechler, 2010), and "MuMIn" (Barton, 2014), (see Appendix C for a brief introduction to R),

RTs (in millisecond) were entered into the model as the dependant variable. The model also included four independent variables: Subject, Item, Item Type, and Language. The data points with standardise residuals of above absolute 2.5 standard deviations were eliminated from the dataset (Baayen, 2008; Baayen & Milin, 2010) and 3.4% of the data was lost. The models were then built using the trimmed data.

In building the model, two estimates of Item and Subject were entered into the model as crossed-random effects as each item was seen once by each participant. Estimates of Item Type (collocation vs. non-collocation) and Language (Persian vs. English) were also entered into the model as two fixed effects. In addition, as the interaction of two fixed effects of Item Type and Language was of interest to us, this interaction was also included into the model.

The model development procedure progressed as follows. The first step was to design a maximal model which included the crossed-random effects of Item and Subject. This model also included random slopes ((1+ItemType|Subject) + (1+Language|Item)) to allow for individual and group variations.

We then used a backward stepwise procedure to remove variables that did not enhance the model fit. The process included elimination of the estimates with lowest t value and then fit the model. This process continued up to the point where only estimates with a t value of at least two remained in the model. At this stage, a full model comparison was carried out using model.sel function in the MuMIn package (Barton, 2014) in R. This function provides estimates of the corrected Akaike information criterion (cAIC), which is the corrected version of the Akaike information criterion (AIC) used to determine the best model that is corrected so it would not be affected by sample size.

The first group of models was built based on the above procedure using RT as the dependant variable. However, prior to model construction, we first had to prepare the data. As the first step in data preparation, responses faster than 300 milliseconds were eliminated as they were assumed to be erroneous voice recognitions. In addition, wrong responses (including mispronunciation or failure to read the item correctly), which were recorded by the researcher during each experiment, were identified and detected. The items that had timed out at 2000 ms. Then, the remaining RTs were using natural logs.

Table 1. Fixed-effects estimates of best-fit model

Random effects	**Variance**	**Standard Deviation (SD)**	**Correlation**
items (Intercept)	0.010220	0.10109	
subjects(Intercept)	0.004521	0.06724	
Language(condition slope)	0.005876	0.07666	-0.93
Residual	0.016018	0.12656	

Fixed effects	Estimate*	Std. Error	Degrees of freedom	t value	Pr(>|t|)
(Intercept)	0.03454	0.03454	48.08000	195.233	< 2e-16
LanguagePersian	-0.21930	0.04649	48.23000	-4.717	2.08e-05

Note. * = Scale of all estimates is in logged milliseconds

Results are summarized in Table 1. The top section of Table 1 provides information on the variability in the data that is related to random effects (e.g., some participants were fast readers, whereas others were slow readers). It is apparent that there is a relatively large amount of variability in reading speed between participants (SD = 0.010220), much less between language slope (SD = 0.005876), and relatively little between items (SD = 0.004521). The "Residuals" row displays information on the residual error (i.e., the amount of variability in the data that was not explained by the model) (SD = 016018).

The bottom part of the table provides information on the fixed effect (i.e. constant effect over subjects and items). The "the intercept" row lists the mean RT for English (-0.03454 on the log scale) and the LanguagePersian row informs us that the difference in read-aloud time between English group and Persian group is -0.21930 ms (on the log scale). The mean RT for Persian group could be calculated by adding this difference to the English mean ((0.03454 ms) + (-0.21930 ms) = 0.18476 ms on log scale). The t- and p-values indicate that the difference between English and Persian groups is significantly different from zero. As RTs were log-transformed to be entered into the model, to obtain RTs in milliseconds again, we applied the exponential function [MEn = exp(0.3454) = 1.41256 ms], [MPer = exp((0.03454)+(-0.21930) = 0.83130 ms].

After accounting for subject and item variability, Persian group [MEn = exp(0. 3454) = 1.41256 ms] were read 1.41 ms slower than English group [MEn = exp((0.03454) + (-0.21930)) = 0.83130 ms ; t(681) = - 4.717, p = 2.08e-05].

As stated previously, the researcher hypothesised that the Language would not affect read-aloud. However, the findings of the read-aloud experiment in this study provided as counterevidence for the hypothesis, suggesting that language affects processing speed. Moreover, the hypothesis on the effect of Item Type was rejected based on p-value of effect in the results.

4. Discussion

Our aim in this study was to investigate whether there is processing advantage for collocations compared to non-collocations for bilinguals and whether the pattern identified was the same for monolinguals. The study perceived its research questions through Wray's dual-route model. The overall aim was to investigate if at high levels of L2 proficiency either of the analytic or holistic processing would be involved.

In sum, results from the read-aloud study suggested that on average, the monolingual group spent less time for completing the task than the bilingual group (disregarding whether the items were collocations or non-collocations). Also, no significant effect was observed for item type on RTs.

The literature on processing of formulaic language has provided more evidence in recent years. However, diversity in experimental features makes the available evidence on this type of processing inadequate for drawing any clear conclusions.

The results of the read-aloud task in the current study revealed no significant effect for item type, as no advantage was observed for non-collocations over collocations, and the results were the same for both groups of participants in either of the tasks. These findings contradict our hypothesis that was based on the assumptions of the theories on holistic retrieval and processing, which predict higher speed of processing for collocations, as they are believed to be represented and restored as single lexical units (cf., Wray, 2002).

The conflicting findings of this study to that of mentioned research could be explained through the differences in their experimental designs. One significant difference is the types of formulaic sequences used as stimuli. Studies reporting processing advantage for formulaic sequences have mostly focused specifically on lexical bundles. Although non-idiomatic, they have many pragmatic functions in discourse, but still are literal in a semantic sense, for example, in the middle of or on the other hand (Tremblay et al., 2011; Biber et al., 1999). Hence, formulaic processing of sequences differs in this sense from the aforementioned studies, as collocations are to a fair degree transparent in terms of meaning, and also to a fair degree restricted in terms of structure (Cowie, 1998).

The noncollocations in the stimuli in the current study are semantically very transparent as their meanings rely on the meaning of individual words. The collocations, constituting adjective-nouns, were also quite transparent. This high degree of semantic transparency for both types of items may have required the same amount of time for reading aloud them aloud by both language groups. This could be supported by the difference observed in Han (2015), where verb-noun collocational items used in their strudy were of semantically less transparent nature in comparison to non-collocational items which were very transparent, and were read faster.

The results may also highly represent the fact that collocations of such type are not processed as single chunks (Tremblay, 2011). It has been greatly argued in the literature that smaller reaction times are indicative of processing and representation of formulaic sequences as a chunk from and in the mental lexicon (see Siyanova-Chanturia (2015) for a discussion). As the results of this study revealed no significant difference between the two type (collocations and non-collocations), it could hence be claimed that adjective-noun collocations with high frequency of use are not processed as a whole as predominantly stated by Wray (2002).

Our second research question examined the difference between RTs' of monolingual and bilinguals in reading aloud collocations. Although in this study the high proficiency level of bilingual group controlled by means of both the self-rated proficiency estimates and the OPT scores, the L1 speakers' processing, when compared to L2 speakers at advanced levels of proficiency, did not show an advantage in production rate of the collocational items through the read-aloud task. Despite being counterevidence to our hypothesis, the results were consistent with the findings of most previous research which indicated that native speakers processed formulaic sequences with higher rates than non-native speakers. The findings were all the same irrespective of the methodological features and methods, or the L2 under examination. Although, the bilingual group of current study only included those at the advanced proficiency levels, their processing efficiency still did not reach the same level as the monolingual group.

The dual-systems processing model (Wray, 2002) may be used to explain the differences observed between L1 and L2 collocational processing in our results. The assumptions of the dual-route model state that direct retrieval is made possible through the formulaic sequences stored mental representation in the mental lexicon. This enhances the speed of the route for comprehension and production, but not for computation which is used for multiword units without a stored mental representation. In addition, in this model both holistic and analytic language processing is developed by a L1 learner and these strategies are believed to be employed again when learning a L2. However, for L2, the analytic processing is believed to have a major role, and hence, be the reason for the disadvantage that L2 learners experience in language processing speed. Looking back at the participants of this study, Persian-English bilinguals in the research could have longer RTs simply because they processed the high frequency adjective-noun collocations through the analytic route. Consequently, it may be concluded from the results that the speakers with high levels of proficiency and experience in a L2, process formulaic sequences analytically in the L2, resulting in a processing cost for L2 in comparison to L1. However, the fact that item type (collocation vs. non-collocation) has no significant effect on processing rate suggests that although this type of collocations may not be processed as single cognitive chunks in both L1 and L2 (as discussed earlier), the rate of difference observed here may well be due to the processing speed difference at single word level between L1 and L2.

All in all, to reach a more definite understanding of collocational processing, there is yet need for more empirical evidence to inspect its other aspects such as congruency and semantic transparency. It may also be a good idea to conduct studies including speakers with L1 and L2s from other language families as participants.

As this study took a frequency-based approach, there is need for further investigations of collocations from the phraseological approach and also other examinations in which more collocational features are studied. However, it should be mentioned that as this study chose high-frequency collocations and congruent collocations to work with it cannot make any claims based on lower frequencies and incongruent collocations. Also, as at higher frequency bands collocations have higher degrees of transparency, the degree of semantic transparency was only controlled, leaving it to be further investigated by variating its different levels.

In the current study, the items and trials were controlled for a number of potential confounding variables such as whole item frequency, average number of letters, average number of adjective letters, average number of noun letters, and average frequency of adjectives and nouns separately. The assumption was that as the collocations were chosen from the same frequency band there would be no significant difference in their processing speed. The results from both experimental tasks, however, did not support the hypothesis, indicating that even at high levels of proficiency there is a processing advantage for one's L1 over their L2. Controlling the above mentioned factors also posits a potential for other factors and characteristics of collocations that may have resulted in such results.

The items in this study were chosen from among the adjective-noun constructions in both Persian and English languages. The motive behind this choice was the amenability of this type of collocations, as this is a word combination that is almost similar in both languages (with the exception of the directions of the two writing systems, and the Ezafe construction in Persian which are inevitable). The item development in this study followed a rigorous procedure in which several factors were controlled, including: whole frequency, noun frequency, adjective frequency, MI scores, t scores, and number of letters. To control these criteria for such research in the future seem to be necessary.

The research makes comparison between the bilingual speakers' formulaic language processing rates with a monolingual group which share the same L1. To the best of our knowledge, to date, all but an unpublished research by Sugiura et al. (2013), have compared L2 speakers of English to an English native group. The downside to comparing a L1 group with a language different from the bilinguals' L1 is that the influence of different features which a speaker's L1 might have on processing of collocations is neglected, as the comparisons have been made between a group of English native speakers and a group of non-native speakers of English which has different features in contrast to languages such as Persian, Chinese, and Arabic.

In addition, in this research we compared a group of Persian monolinguals to Persian-English bilinguals, which only knew one language and two languages respectively (as declared in a self-report). So, the L1 participants were only selected amongst those who were monolinguals of Persian and spoke no other languages. This was with the aim of eliminating any effects that could be caused by other languages that a speaker might have known.

This study also involves pedagogical implications for the current research in the realm of L2 learning. The findings of this study supported the suggestion about the advantage of item type on language processing, which would mean that teaching collocations and formulaic units would in fact have positive effects on learning. This could result in higher rates of learning when dealing with learners' vocabulary skills. Formulaic language units are also suggested to enhance learners' fluency and ensure a more native like speech. Hence, these language constructions require more attention from the syllabus designers and teachers in comparison to single-word units. This need, as advocated by our findings in this study, exists even at high levels of L2 proficiency.

There are areas that future research in the realm of formulaic language processing may focus on in future. The present study only focused on congruent English and Persian collocations. Further research is still needed to identify if the results of this research will be repeated for the incongruent collocations (i.e. those with no literal equivalent in the L2), as well. Also, as this study merely included high-frequency collocations, which were compared to non-collocations, the differences between other frequency bands still remain to be examined further, as this can elaborate upon how collocations are processed by bilingual speakers, especially bilinguals with different L1s. Investigation of L1 and L2 in different participants allowed us to investigate the degree with which monolinguals and bilinguals differed in collocational processing of a first and a second language. However, future research on this issue could help identify if collocational processing of L1 is altered in bilingual speakers in comparison to monolinguals. In other words, future research could consider comparing bilinguals in both their L1 and L2 and compare the processing patterns to a monolingual group.

This research, examined collocational processing using a productive online experimental tasks. Future investigation into receptive mode of processing may be beneficial to be carried out as they can add more to our knowledge and this can help to identify how the processing of collocations and non-collocations differ when dealing with different modalities.

In addition, as the number of participants may be an influencing factor on the data analysis of RT studies, further studies with larger number of participants could elevate our understanding of the effects of this factor. Also, if further advances in the Persian corpora are achieved, the number of items may also be increased which could also prove to be beneficial for improvements in the results.

5. Conclusion

It could be argued that the literature, to date, has emphasized on the role of formulaic sequences, although with less attention on collocations, specifically in the area of language processing. The present study was interested in the role of formulaic language units and whether they affect how L1 and L2 are processed. It also considered whether at advanced levels, language learners' Collocational processing followed a similar pattern to that of a monolingual's L1 processing. This research examined these assumptions using a read-aloud task. The results supported previous findings in the literature. The findings of this research can add to the existing knowledge on the processing directions and feature of monolinguals and bilinguals. However, to attain a clearer vision and understanding, the literature on the processing of formulaic sequences requires yet more research to be done.

References

Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon* (4th Edition). John Wiley & Sons.

Arnon, I., & Priva, U. C. (2013). More than Words: The Effect of Multi-Word Frequency and Constituency on Phonetic Duration. *Language and Speech*, *56*(3), 349–371.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82.

Baayen, H. (2008). *Analyzing linguistic data*. *A Practical Introduction to Statistics Using R. Cambridge, UK: Cambridge University Press*.

Baayen, H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28.

Barfield, A., & Gyllstad, H. (2009). Introduction: Researching L2 collocation knowledge and development*. In A. Barfi eld & H. Gyllstad (Eds.), Researching collocations in another language: Multiple interpretations (pp. 1 – 18).* Basingstoke, UK: Palgrave Macmillan

Biber, D., & Conrad, S. (1999). Lexical bundles in con- versation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 181–190). Amster- dam/Atlanta, GA: Rodopi.

Bijankhan, M., Sheykhzadegan, J., Bahrani, M., & Ghayoomi, M. (2011). Lessons from building a Persian written corpus: Peykare. *Language Resources and Evaluation, 45,* 143–164.

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, *27*(6), 668–683.

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.

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, *32*, 45–61.

Cowie, A. P. (Ed.) (1998). Phraseology: Theory, analysis and applications. Oxford, UK: Clarendon Press.

Cutting, J. C., & Bock, K. (1997). That’s the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition*, *25*(1), 57–71.

Durrant, P. L. (2010). *High frequency collocations and second language learning*. Nottingham: University of Nottingham PhD dissertation.

Ellis, N. C., & Simpson-Vlach, R. (2009). Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory*, *5*(1), 61–78.

Gyllstad, H. (2007). *Testing English collocations Developing receptive tests for use with advanced Swedish learners* (Unpublished doctoral dissertation). Lund University, Sweden.

Gyllstad, H., & Wolter, B. (2015). Collocational processing in the light of a phraseological continuum model: Does semantic transparency matter? *Language Learning*, *66*(2), 296–323.

Han, S. (2015). *Processing formulaic sequences by native and nonnative speakers of English: Evidence from reading aloud* (Unpublished doctoral dissertation). Northern Arizona University, Arizona.

Janssen, N., & Barber, H. a. (2012). Phrase frequency effects in language production. *PloS One*, *7*(3), e33202. Doi: 10.1371/journal.phone.0033202

Lehmann, H.-M., Hoffmann, S., & Schneider, P. (2002). BNCweb: A webbased interface to the British National Corpus. Version 2.0. Zurich: Zurich University.

Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*, *32*(2), 129–148.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication*, (pp. 191-225). London: Longman.

Schmitt, N. (Ed.). (2004). *Formulaic sequences: acquisition, processing and use* (Vol. 9). Amsterdam: John Benjamins Publishing.

Shaoul, C. (2012). *The Processing of Lexical Sequences*. University of Alberta.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, *27*(2), 251–272.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: the role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 776–784.

Siyanova, A., & Schmitt, N. (2008). L2 Learner Production and Processing of Collocation: A Multi-Study Perspective. *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes*, *64*(3), 429–458.

Sosa, A. V., & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word of. *Brain and Language*, *83*(2), 227–236.

Sugiura, M., Fujumura, I., Yamashita, J., & Leung, C.-Y. (2013). Native-like Advantage in Processing Collocational Expressions among Japanese Learners of French

[PowerPoint slides]. *Euro SLA*. Retrieved from http://sgr.gsid.nagoyau.ac.jp/wordpress/wpcontent/uploads/2014/02/

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, *61*(2), 569–613.

Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics*, *16*(2), 180-205.

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.

Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing. *Studies in Second Language Acquisition*, *35*(3), 451–482.

Wolter, B., & Yamashita, J. (2014). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, *1*(29), 1–29.

Wray, A. (2002). *Formulaic language and the lexicon* (Vol. 5). Cambridge University Press Cambridge.

Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, *44*(4), 647–668.

Chabahar Maritime University
Iranian Journal of English for Academic Purposes ISSN: 2476-3187
IJEAP, (2016) vol. 5 issue. 1 (Previously Published under the title: Maritime English Journal)

Appendices

Appendix A

Experimental items

Table 2. List of collocational items

	High frequency English collocations	High frequency Persian collocations
1	other side	طرف دیگر
2	twentieth century	قرن بیستم
3	human rights	حقوق بشر
4	private sector	بخش خصوصی
5	foreign affairs	امور خارجه
6	second half	نیمه دوم
7	large number	تعداد زیادی
8	recent years	سال‌های اخیر

Table 3. List of non-collocational items

	English non-collocations	Persian non-collocations
1	young pieces	قطعات جوان
2	normal walls	دیوار طبیعی
3	common call	ندای مشترک
4	current schools	مدارس آفتابی
5	easy rest	استراحت آسان
6	first fields	زمینه اول
7	clear player	بازیکن آشکار
8	low generations	نسلهای پایین

Appendix B

Linear mixed-effects modelling

Linear mixed-effects modeling (LMM) is a state-of-the-art technique increasingly being used for data analysis in language-related studies. As the name suggests, LMMs contain both random effects of subject and item and fixed effects of independent variables. In LMMs random effects indicate variables sampled from a large population, for example, items or participants. So they are different from one experiment to another, and are also not of interest in the research. Once the participants see items as in this study, subject and item random effects are crossed. Thus, fixed effects are independent variables whose levels indicate the entirty population, such as, language group (e.g., monolingual vs. bilingual) and certain levels of fixed effects are also of interest in research.

Analysis of Variance (ANOVA) and t-tests have been applied extensively in language research, however, more grouped data research roused more interest in a class of statistical models, particularly mixed-effects models in general, or more specifically linear mixed-effects models (LMMs) (Baayen, 2008).

Appendix C

R statistical programme

R is an open-source statistical, which is free to download from http://www.r-project.org (in this study we used the latest version available at the time (version 3.2.4.)). Although those researchers trained in statistical software such as SPSS, which provide a graphical interface, may find it daunting to work with R as it is a command-line rather than menu driven at first, it will become a highly flexible tool after a while. As it is beyond the scope of this article we will not provide an intoduction to R syntax, however those interested may refer to textbooks such as Baayen (2008), Crawley (2007), Dalgaard (2008) or Vasishth and Brow (2011).

Through the standard installation of R a variety of statistical analyses can be performed, and additional packages which offer tools specialized for different types of statistical analyses. The packages of interest in this research are the lme4 (Bates & Maechler, 2010), which provides state-of-the-art tools for running mixed-effects models, and MuMIn (Barton, 2014), which is used for model selection and obtaining a robust *p*-value.

Chabahar Maritime University
Iranian Journal of English for Academic Purposes ISSN: 2476-3187
IJEAP, (2016) vol. 5 issue. 1 (Previously Published under the title: Maritime English Journal)

Parastoo Harati Mokhtari is an MA graduate of Applied Linguistics from Tarbiat Modares University, Tehran, Iran, and holds a BA in English Translation from Chabahar Maritime University, Chabahar, Iran. She has finished all level of her education with an Exceptionally Talented Students' Award. Her research interests include Psycholinguistics (language processing and representation, specifically on formulaic language), Second/Foreign Language Acquisition, and Bilingualism.

Reza Ghafar Samar is an associate professor of Sociolinguistics and Applied Linguistics at Tarbiat Modares University. He holds a Ph.D. from University of Ottawa, Canada in Applied-Sociolinguistics, an MA in TEFL from Tarbiat Modares University and a BA in English Language and Literature from Ferdowsi University. He specializes in Sociolinguistics (Variation Theory, Bilingualism, and Language change) and its application in Second/Foreign Language Acquisition. He is the author of numerous books like "Aspects of Second Language Speech: a variationist perspective" published in 2010 and articles appearing as book chapters or in journals such as International Journal of Bilingualism.

Gholam Reza Kiany received his PhD degree from Essex University in the UK. He is currently an associate professor in Tarbiat Modares University. After his years of teaching different courses including research methodology, language testing, and psycholinguistics, he has taken perspectives on CAT, program policy and evaluation as his recent areas of interest. He has authored several English books and EFL-related papers. He has also conducted a number of national ELT research projects in the Ministry of Education and Ministry of Higher Education.