

REPLICATION STUDIES

Research on the on-line processing of collocation: Replication of Wolter and Gyllstad (2011) and Millar (2011)

Suhad Sonbul¹ and Anna Siyanova-Chanturia^{2*}

¹Umm Al-Qura University, Makkah, Saudi Arabia and ²Victoria University of Wellington, Wellington, New Zealand

*Corresponding author. Email: anna.siyanova@vuw.ac.nz

Research employing psycholinguistic techniques to assess the on-line processing of collocation by native and non-native speakers has flourished in the past few years. This line of research aims (among other things) at exploring actual performance in real time as opposed to the traditional paper-and-pencil testing techniques that have been extensively employed in collocation research. The present paper reviews some of the pertinent research on the on-line processing of collocations and argues for the need for more replication studies in the area. It then looks at how two experimental studies on the topic – Millar (2011) and Wolter and Gyllstad (2011) – may be replicated in order to gain deeper understanding of the key factors behind collocation processing and to obtain more valid and generalizable results that can find their way into language teaching practice.

1. Introduction

Interest in the processing of formulaic language, idioms in particular, has its roots in the 1980's psychology literature (see, for example, Gibbs & Gonzales, 1985; Gibbs, Nayak, & Cutting, 1989). Recent years, however, have seen a growing interest in the topic with respect to other types of multi-word sequences, such as collocations, binomials, and lexical bundles. This is, to a great extent, due to the observation that formulaic language is a pervasive phenomenon in speech and writing, serving a wide range of pragmatic, discourse, and cognitive functions (for an overview of formulaic language from a cognitive, pragmatic, and pedagogical perspective, see Siyanova-Chanturia & Pellicer-Sanchez, 2018).

While researchers have long advocated the use of psycholinguistic techniques in assessing on-line language processing in a first (L1) and second (L2) language (Juffs, 2001; Marinis, 2003), employing such techniques in the context of formulaic language – above and beyond idiomatic expressions – is a relatively recent phenomenon (for an overview, see Siyanova-Chanturia & Van Lancker Sidtis, 2018). The present paper focuses on one type of formulaic language, collocations, and their on-line (real-time) processing as opposed to their off-line usage, as seen in traditional paper-and-pencil tests.

In the area of collocations, studies involving off-line measures have shown that L2 learners' knowledge of collocation often lags behind that of L1 speakers. This pertains to the research on error analysis (Granger, 1998; Nesselhauf, 2005; Laufer & Waldman, 2011) and paper-and-pencil assessment techniques (Bahns & Eldaw, 1993; Nguyen & Webb, 2017). Much of this research has shown that learners' L1 can have a considerable influence on the knowledge of L2 collocation.

Recently, a number of studies investigated the on-line processing of collocation with several pedagogical implications in mind. One such line of research, for example, employed various reaction-time techniques to examine the effect of L1 on L2 collocation processing (Yamashita & Jiang, 2010; Wolter & Gyllstad, 2011; Wolter & Yamashita, 2015, 2018), as well as the interaction between corpus-derived frequency and L1 influence on L2 processing (Wolter & Gyllstad, 2013). Another line of enquiry, which was spearheaded by Millar (2011), has focused on using psycholinguistic techniques to

complement traditional error analysis methods. The studies cited above, although innovative and important for the field of second language acquisition, have been noted to suffer from several methodological limitations, or have, indeed, produced contrasting results. This necessitates conducting replication studies to arrive at more valid, reliable, and generalizable findings.

The present paper suggests replications of two empirical studies from within this literature – Millar (2011) and Wolter and Gyllstad (2011). We begin with a brief overview of the relevant literature, followed by an analysis of each study and suggestions for possible replications. The final section presents concluding remarks and directions for future research.

2. Background

Although the way collocations are treated varies depending on the approach adopted (e.g., phraseological versus frequency-based), researchers agree that these phrasal configurations are not as fixed as other types of formulaic sequences (such as idioms and proverbs). For example, Howarth (1998) views free combinations, collocations, and idioms as forming a continuum, proposing that most learners' difficulties are related to collocations (the central point on the scale). For the purpose of the present paper, collocations are defined as two-word combinations (bigrams) with above-chance corpus frequency, as evident in measures of association strength, such as t-score and mutual information/MI (see Gablasova, Brezina, & McEnery, 2017 for an overview). Research examining L2 collocation performance can be broadly divided into two categories: (1) corpus-based research analysing English as a foreign/second language (EFL/ESL) learners' written or, more rarely, spoken output and (2) research employing off-line and on-line measures to explore issues specific to collocation processing and use.

Studies conducted under the first category (e.g., Granger, 1998; Nesselhauf, 2005; Laufer & Waldman, 2011; for a review, see Granger, 2018) have generally reached the conclusion that EFL learners (L1 French, Hebrew, and German, respectively) use fewer typical, native-like collocations compared to L2 speakers (but see Siyanova & Schmitt, 2008, Study 1 for different results), with a clear influence from their L1. In classifying learners' errors, these studies have mostly used L1 judgements (e.g., Granger, 1998), but, more recently, other external measures have been employed, such as idiomatic dictionaries (e.g., Nesselhauf, 2005) and corpus-based norms (e.g., Siyanova & Schmitt, 2008; Laufer & Waldman, 2011). In general, studies on error analysis have often been criticized for using such off-line measures in error identification and classification (e.g., Ellis & Barkhuizen, 2005). In an innovative study, Millar (2011, see below for a detailed review) combined error analysis with a self-paced reading task borrowed from experimental psychology to provide a more objective measure of how L2 collocation errors, atypical of L1 usage, are treated and processed on-line.

Another area of research on collocation performance is concerned with assessing L2 knowledge of collocations using paper-and-pencil tests. Studies employing measures of this type found that EFL learners' off-line collocation performance was particularly weak ranging between 53% in a translation task and 48% in a cloze task for German learners (Bahns & Eldaw, 1993), 50% in a multiple-choice test for Vietnamese learners (Nguyen & Webb, 2017), and 56% in a cued recall task for Spanish learners (González Fernández & Schmitt, 2015).

Siyanova & Schmitt (2008, Study 2 and Study 3) were probably the first to combine traditional off-line tests with on-line measures to test both L1 and L2 speakers. The results of a timed rating task used in their study showed that L2 learners only exhibited native-like sensitivity to corpus-derived frequencies after spending an extended period of time in the ESL context (see also Sonbul, 2015 for similar evidence from an eye-tracking experiment). Phrase frequency effects have since been reported in a wealth of studies, in particular in the context of usage-based approaches to language acquisition, processing, and use (for a review, see Siyanova-Chanturia & Van Lancker Sidtis, 2018; Wulff, 2018).

Another prominent line of research is the effect of L1 on L2 formulaic language processing (for an overview, see Conklin & Carrol, 2018). With respect to collocations, Yamashita and Jiang (2010) used a timed judgement task (yes/no acceptability judgements) to examine this issue. The study included both congruent ($L1 = L2$) and incongruent ($L1 \neq L2$) collocations and tested two groups of Japanese

learners (EFL intermediate and ESL advanced), as well as a control group of L1 speakers. Only the lower-level EFL group showed a processing advantage for congruent collocations over incongruent ones (in the by-participant analysis only). It was concluded that as L2 proficiency increases, the L1 effect fades away. This line of research is important as it can shed light on the organization of the L2 mental lexicon (e.g., Kroll & Stewart, 1994), as well as inform teachers and material developers as to the relative ease/difficulty of collocations with/without L1 equivalents.

The pioneering study by Yamashita and Jiang (2010) has since been followed by a series of follow-up studies by Brent Wolter and his colleagues (see Table 1). As can be seen in the table, given the variety of tasks, items, subjective proficiency levels, analyses performed, and the degree to which L1s in question differ from L2 English (i.e., Japanese and Swedish), it is difficult to reach a clear conclusion regarding the congruency effect in L2 collocation processing. Each of these studies appears to be adding one piece to the puzzle without actually solving it completely. We believe that this line of research may, in fact, represent a classic example of what Mackey (2012, p. 29) considers ‘ripe for replication’, with inconsistent findings in need of further investigation.

As is evident in this short (and inevitably selective) review, pedagogically oriented research employing on-line measures to explore L2 collocation processing, though promising, is still in its infancy and is characterized by a number of methodological limitations that are likely to affect the validity and generalizability of the results reported. The scarcity of this research and the limitations associated with it point to the need for replication studies that can systematically deal with the limitations in the original designs (approximate replications) and re-explore the issues using different methods and techniques (conceptual replications).

The present paper looks at two studies cited above – Millar (2011) and Wolter and Gyllstad (2011) – and proposes several ways in which they can be replicated. The studies in question used psycholinguistic techniques (self-paced reading and priming, respectively) in assessing on-line collocation processing and were published in *Applied Linguistics*, thus, showing a clear link between the theoretical issues explored and actual practice. Finally, the appendix to both studies contains a full list of items and materials, thereby facilitating replication endeavours.

3. The original studies and approaches to replication

3.1 Wolter and Gyllstad (2011)

As indicated above, the effect of L1 on L2 collocation processing has been studied in some detail. This study by Brent Wolter and Henrik Gyllstad is one of the first and most-cited in this area. The study is also interesting in that it includes both off-line data from paper-and-pencil tests and on-line data from a lexical decision task (LDT) within the priming paradigm, allowing for some useful comparisons. The assumption behind such a priming task is that reaction times (RTs) to a target (the second word within a two-word collocation) will be faster following a related prime (collocate) than an unrelated one (control).

Wolter and Gyllstad (2011) conducted an experiment with two groups of adult participants: 35 native English speakers (American university students) and 30 advanced (based on self-reported scores) Swedish EFL learners. The study included three types of Verb + Noun (object) pairs: congruent collocations (33), incongruent collocations (33), and control non-collocate pairs (33). The authors tested three hypotheses: (1) L2 learners will show a processing advantage for congruent collocations over control pairs, (2) L2 learners will show a weaker advantage for incongruent collocations than congruent collocations (in comparison to control pairs), and (3) L1 speakers will not show any differences between the processing advantage for congruent and incongruent collocations over control pairs.

The target collocations were extracted from the British National Corpus (BNC) with a minimum frequency of 5 and a minimum t-score of 2.0. There were no significant differences between congruent and incongruent collocations in terms of their t-score or raw BNC frequency. Moreover, the three sets (congruent, incongruent, and control) were controlled for the length, frequency, and concreteness of

Table 1. Summary of studies investigating L1 effect on L2 collocation processing

Study	Task	L2 learners (in addition to L1 baseline)	Items	Analysis	Summary of results for L2 learners
Wolter & Gyllstad (2011)	Single lexical decision task (priming)	Advanced EFL learners, L1 Swedish	Congruent, incongruent, and baseline	MANOVA (F1 and F2 analyses)	Congruency effect (in F1 analysis)
Wolter & Gyllstad (2013)	Acceptability judgement task	Advanced EFL learners, L1 Swedish	Congruent, incongruent, and baseline	Mixed-effects modelling	Congruency effect
Wolter & Yamashita (2015)	Double lexical decision task	Intermediate EFL learners, L1 Japanese	Incongruent, L1-only, and baseline	Mixed-effects modelling	No processing advantage for incongruent or for L1-only collocations over controls
Wolter & Yamashita (2018)	Acceptability judgement task	Intermediate and advanced EFL learners, L1 Japanese	Congruent, incongruent, L1-only, and baseline	Mixed-effects modelling and correlations	Congruency effect for both groups and no processing advantage for L1-only collocations over controls

individual words (which included cognates). A norming collocation familiarity-rating task was administered to 20 L1 speakers prior to the study confirming condition status with significant differences between the baseline control condition and the other two conditions. It should be noted here that little is reported in the study regarding the selection of non-collocate pairs. However, looking at the items provided in the appendix, one can clearly observe semantic implausibility of most control items (e.g., **tell rug*, **invite success*). This alone is likely to cause a significant delay in the processing of such phrases. Of note is that this also applies to the other studies summarized in Table 1.

In the priming LDT experiment, the prime verb (first word of the pair) was presented on a computer screen for 250 ms followed by a blank screen for 50 ms; then the target noun (second word of the pair) appeared and a lexical decision (whether or not the word on the screen is a real English word) was required through pressing Yes/No-assigned keys on a standard keyboard. The target items were counterbalanced across three experimental lists so that no prime or target was repeated more than once, and the presentation of items was randomized. Response latency to and decision accuracy on the target nouns were recorded using the DMDX software (Forster & Forster, 2003). In addition to the experimental items, 440 filler items were included in the experiment comprising either real word targets (121) or non-word targets (220). In addition to the priming experiment, L2 learners were also administered the Yes/No COLLMATCH collocation test (Gyllstad, 2007) to assess their receptive knowledge of the same target collocations.

The analysis was done using traditional MANOVA (F1 and F2) analyses. The analysis of L1 data showed a clear processing advantage (faster RTs) for both congruent and incongruent collocations relative to control pairs but no difference between the two types of collocations (this was observed in both F1 and F2 analyses). On the contrary, L2 learners exhibited a processing advantage for congruent collocations over control pairs (in both F1 and F2 analyses), but the difference between incongruent collocations and the other two types of items (congruent collocations and control pairs) was only significant in the by-participant analysis (F1). The same pattern of results was observed when cognates were excluded from the analyses. The results of the off-line test showed more 'Yes' responses to congruent collocations than to incongruent collocations. On the basis of these results, Wolter & Gyllstad (2011) concluded that learners' L1 has a strong influence on how EFL learners process collocations in their L2 (although the effect appeared to be stronger off-line than on-line).

3.1.1 Approach to replication

Given the discrepancy between the reported findings on congruency effects in L2 collocation processing (see Table 1), Wolter and Gyllstad (2011) can be approximately replicated in a series of studies, wherein each study controls for one of the experimental variables to uncover its potential effect on L2 processing.

A starting point, for instance, can be a study involving the same priming experiment and the same experimental items as in the original study but including three different groups of participants: (1) native speakers of English (same as original), (2) advanced Swedish EFL learners (same as original), and (3) intermediate-level Swedish EFL learners (a new group). A more objective estimate of L2 proficiency (e.g., vocabulary size test scores) can be used instead of self-reported proficiency scores used in the original study. The analysis can include a MANOVA to measure whether the original on-line congruency advantage (evident in the F1 analysis only) can be stronger for a lower proficiency group. Although two of the studies exploring the collocation congruency effect in L1 Japanese learners included two levels of proficiency (Yamashita & Jiang, 2010; Wolter & Yamashita, 2018), the two studies involving L1 Swedish only included advanced learners (Wolter & Gyllstad, 2011, 2013) (see Table 1). It will, thus, be interesting to explore the effect of proficiency on the congruency effect in EFL Swedish learners.

The proposed replication study can benefit from a more complex and advanced statistical analysis, such as mixed-effects modelling which can include both participants and items as random effects in one model. Mixed models can also accommodate a range of continuous variables that are otherwise difficult to include in traditional analyses (see Baayen, Davidson, & Davidson, 2008 for advantages of

mixed-effect modelling). A mixed-effects model can be fitted including the main effects (group and item type) in addition to proficiency scores as fixed factors and participants and items as random factors. This analysis can add several pieces of evidence to our understanding of L1 congruency effect and deal with the limitations related to the ways in which the analysis was conducted in the original study (with separate F1 and F2 analyses).

Another approximate replication study can control for the potential effect of item type in a series of experiments. The first experiment in this replication study can employ exactly the same task (priming and LDT), two groups of participants as in the original study (native speakers and advanced Swedish EFL learners), and the same experimental items (congruent and incongruent collocations). The difference, and, in our opinion an improvement, will be the way in which control (baseline) items are devised. As reported above, most control items in the original study were not semantically plausible, which was likely to introduce a confound affecting the results of the study. In addition, not much is stated in terms of the corpus-derived frequency and association strength of control items in the original study. Control items for this replication experiment can be devised in a way that makes them semantically plausible but with very low (close to zero) frequency and low association (e.g., **extreme mistake*, **impressive explanation*, see Sonbul, 2015). In a norming study, a group of native speakers can be asked to rate control items for ‘semantic plausibility’ in order to include only those that are plausible. If the pattern of results of the main experiment for L1 and L2 speakers is similar to what was found in the original study, this can be taken to suggest that semantic plausibility is not a confounding factor. If, on the contrary, the results of the replication study diverge from the original findings, this will suggest that greater care should be exercised when creating control items in similar experiments.

Together, these multi-step replication studies can deepen our understanding of the effects of L1 on L2 collocation processing, improving both the internal validity and the ecological validity of the study.

3.2 Millar (2011)

As discussed above (see Section 2), research on learner corpora (e.g., Nesselhauf, 2005) has mainly used native speakers’ off-line judgements (intuitions) to identify learner errors in the use of L2 collocation. In his study, Neil Millar employed an innovative design to explore how L1 speakers process learner collocation errors in real time. The main aim of the study was to explore the widely held assumption that the presence of atypical collocations in L2 production constitutes a hindrance to communication.

Millar (2011) compiled a corpus of English essays collected from 960 EFL Japanese university students (approximately 180,000 tokens). From this corpus, atypical significant two-word pairs were extracted. These items had a minimum frequency of 2 (in the learner corpus), were produced by a minimum of two learners, were not attested in the BNC, and were judged by the researcher to be non-target-like. For each of the 32 atypical pairs, a semantically related, acceptable collocation was identified (based on its MI score) in which either the first word (23 items) or the second word (nine items) was replaced (but never both). Experimental items were categorized as representing either a ‘lexical misselection’ (25 items) involving wrong word choice (e.g., **best partner* versus *ideal partner*) or ‘misformation’ (seven items) involving a morphological mistake (e.g., **culture background* versus *cultural background*) (p. 137). Of note is that various factors related to component words, including length, frequency, and part of speech, were not controlled for.

Each typical-atypical pair was then inserted into an identical sentence context, created by the researcher based on corpus examples, representing two experimental conditions (atypical and typical). Sentence contexts were rated by 20 native speakers for acceptability, with a significant difference found between the two experimental conditions (except for one item not specified by the researcher). In addition to the experimental items, 16 filler items – all typical and native-like – were included in the study. Each experimental block included 16 typical collocations, 16 atypical pairs, and 16 filler (typical) items. It is noteworthy that atypical word pairs might have thus stood out due to the ratio of 1:2.

The author employed a self-paced reading task where 30 L1 speakers of English read one version of the sentence word-by-word (in two counterbalanced lists with a pseudo-randomized order) by means of clicking an external mouse button using the software PsyScript (<http://lancaster.ac.uk/psychology/research/research-software/>). Following the experimental session, participants answered simple comprehension questions. It should be noted that the instructions provided did not specify whether the participants were required to read aloud or silently. As a result, as Millar (2011) reports, some participants read the sentences aloud, while others read them silently. In addition, many individual words, both in typical and atypical pairs, were repeated in more than one experimental item (e.g., *happiness*, *big*) or in the sentence contexts (e.g., *older*, *people*), which is likely to have led to an unwanted repetition effect.

The analysis (linear mixed models in SPSS) focused on RTs on the second word of the pair (target word), the word immediately following the second word of the pair (+1), and the following word (+2). Results showed that L1 speakers read the second word of typical collocations along with the +1 word and the +2 word significantly faster than atypical (learner) pairs. This result was obtained for the 23 experimental items where the first word was atypical of L1 usage. In a separate analysis, Millar (2011) explored the effect of the error type ('lexical misselection' and 'misformation') on RT data. Only 'lexical misselection' items showed a significant difference in RTs to the target words between the typical and atypical conditions (no analysis was reported for the +1 or +2 positions). Although Millar (2011) argues that it is the small number of items that prevented a significant difference to emerge for the 'misformation' category, one cannot exclude the possibility that L1 speakers might be more accepting of morphological mistakes than those involving wrong word choice.

Based on the reported results, the author concluded that, in line with the widely held belief, L1 speakers do experience difficulty in reading when confronted with an atypical collocation with a 'spill-over effect' from the target atypical item to the following context (Millar, 2011: 142). Given its innovative design and a number of limitations, Millar's (2011) study can be viewed as a fruitful avenue for both approximate and conceptual replications, promoting the use of psycholinguistic techniques in corpus-based research into L2 collocation errors.

3.2.1 Approach to replication

An approximate replication study can involve a more rigorous approach to item selection. As discussed above, the original study included items that differed in a number of ways: (1) classification of deviation as 'misformation' or 'lexical misselection', (2) position of the deviation word (first or second), (3) length in characters, corpus frequency, and part of speech of individual words, and (4) unwanted repetition of individual words (across items and contexts).

An approximate replication of Millar (2011) can focus on a sub-set of more carefully selected items and modify sentence contexts so as to avoid repetition. Including a larger participant pool can help compensate for a small item set. Since length and frequency of the individual words (within each pair) were not experimentally controlled for, mixed-effects modelling can be used to explore the role of these factors (by including them as covariates in the model). If such a replication shows a comparable processing advantage for typical collocations over atypical pairs, this will provide support to the results of the original study. However, if a different pattern of results is observed, this will provide evidence for the need of a more stringent approach to experimental design and items selection.

A conceptual replication is also possible based on Durrant & Siyanova-Chanturia's (2015) evaluation of Millar (2011). These authors reviewed the study in question in some detail, pointing out a number of limitations. First, Durrant and Siyanova-Chanturia (2015) are critical of Millar's (2011) use of the one-word-at-a-time reading mode in a study dealing with multi-word sequences. Having to read multi-word sequences word-by-word is not ideal, as it may encourage single-word processing at the expense of chunking. An alternative is to present target sentences phrase-by-phrase, or, indeed, to use eye movements. Second, the native-like sentence context might put typical (native-like) collocations at an advantage over atypical (learner) pairs. Given that the study attempted to explore how L1

speakers process learner collocation errors, it will be more valid to use authentic (learner-created) contexts.

Thus, the suggested conceptual replication can address the issue of L1 speaker processing of erroneous (learner) collocation using a different methodology, such as a phrase-by-phrase (two words at a time) reading task or eye movements with authentic contexts extracted from the learner corpus. These changes can make a tangible difference, rendering the study more valid by tapping into the processes under investigation. This in turn is likely to increase ecological (pedagogical) validity and thus inform writing assessment practices.

Ecological validity can be further enhanced through a second conceptual replication involving a mixed-methods approach. The quantitative RTs data from a self-paced phrase reading task or an eye-tracking experiment can be fruitfully combined with qualitative data from a follow-up think-aloud protocol task. In this task, L1 participants can be presented with the target sentence contexts off-line and asked to report on their thoughts as they identify and correct collocation errors. These qualitative data, coupled with on-line data, can help us gain a better understanding of how L1 speakers treat and process L2 learners' atypical word pairs.

4. Conclusion

This paper focused on replication suggestions for two studies that are both influential and innovative in the area of on-line collocation processing (Millar, 2011; Wolter & Gyllstad, 2011). We have shown how replication research (both approximate and conceptual) can help solve several methodological issues through a more stringent approach to experimental design and stimuli, the use of advanced statistical analysis, and by means of exploring pertinent research questions from different perspectives. We further advocate combining traditional off-line tests with on-line measures in future research on collocation processing in order to help bridge the gap between experimental research findings and teaching practices in the language classroom.

References

- Baayen, R. H., Davidson, D. J., & Davidson, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21(1), 101–114.
- Conklin, K., & Carrol, G. (2018). First language influence on the processing of formulaic language in a second language. In Anna Siyanova-Chanturia & Ana Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 62–77).
- Durrant, P., & Siyanova-Chanturia, A. (2015). Learner corpora and psycholinguistics. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 57–78). Cambridge: Cambridge University Press.
- Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language*. Oxford: Oxford University Press.
- Forster, K., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments and Computers*, 35(1), 116–124.
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67(S1), 155–179.
- Gibbs, R. W., & Gonzales, G. P. (1985). Syntactic frozenness in processing and remembering idioms. *Cognition*, 20(3), 243–259.
- Gibbs, R. W., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language*, 28(5), 576–593.
- González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure. *ITL-International Journal of Applied Linguistics*, 166(1), 94–126.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 145–160). Oxford: Clarendon Press.
- Granger, S. (2018). Formulaic sequences in learner corpora: Collocation and lexical bundles. In A. Siyanova-Chanturia & Pellicer-Sanchez, A. (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 228–246). Routledge.
- Gyllstad, H. (2007). *Testing English collocations* (Unpublished doctoral dissertation). Lund University, Lund.
- Howarth, A. P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1), 24–44.

- Juffs, A. (2001). Psycholinguistically oriented second language research. *Annual Review of Applied Linguistics*, 21(1), 207–220.
- Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33(2), 149–174.
- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61(2), 647–672.
- Mackey, A. (2012). Why (or why not), when and how to replicate research? In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 21–46). Cambridge: Cambridge University Press.
- Marinis, T. (2003). Psycholinguistic techniques in second language acquisition research. *Second Language Research*, 19(2), 144–161.
- Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*, 32(2), 129–148.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins.
- Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, 21(3), 298–320.
- Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (Eds.). (2018). *Understanding formulaic language: A second language acquisition perspective*. London, New York: Routledge.
- Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, 64(3), 429–458.
- Siyanova-Chanturia, A., & Van Lancker Sids, D. (2018) What on-line processing tells us about formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), pp. 38–61.
- 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.
- 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: A comparison of incongruent and incongruent collocations. *Studies in Second Language Acquisition*, 35(3), 451–482.
- Wolter, B., & Yamashita, J. (2015). Processing collocations in a second language: A case of first language activation. *Applied Psycholinguistics*, 36(5), 1193–1221.
- Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for L2 performance?. *Studies in Second Language Acquisition*, 40(2), 395–416.
- Wulff, S. (2018). Acquisition of formulaic language from a usage-based perspective. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 19–37).
- Yamashita, J., & Jiang, N. A. 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.

Suhad Sonbul is Assistant Professor in the English Language Centre at Umm Al-Qura University. Her research interests include vocabulary learning, formulaic language, and psycholinguistic measures. She has published in several refereed journals including *Language Learning* and *Bilingualism: Language and Cognition*.

Anna Siyanova-Chanturia is Senior Lecturer in Applied Linguistics at Victoria University of Wellington. Anna's research looks at the acquisition, processing and use of vocabulary and multi-word expressions. Anna has published in *Applied Linguistics*, *Language Learning*, *Studies in Second Language Acquisition*, *Journal of Experimental Psychology: Learning, Memory and Cognition*, *Brain and Language*, and other journals.