Productive Collocation Knowledge at Advanced CEFR Levels: Evidence from the Development of a Test for Advanced L2 French

Fanny Forsberg Lundell, Christina Lindqvist, Amanda Edmonds

The Canadian Modern Language Review / La revue canadienne des langues vivantes, Volume 74, Number 4, November/novembre 2018, pp. 627-649 (Article)

Published by University of Toronto Press

Productive Collocation Knowledge at Advanced CEFR Levels: Evidence from the Development of a Test for Advanced L2 French

Fanny Forsberg Lundell, Christina Lindqvist, and Amanda Edmonds

Résumé : Les auteures se penchent sur la relation entre les connaissances productives liées aux collocations et les niveaux élevés de l'échelle du Cadre européen commun de référence pour les langues (CECR). Elles étudient plus précisément le potentiel de progression du niveau B2 au niveau C1 du CECR grâce à l'élaboration d'une épreuve de collocation productive en français langue seconde. Cette épreuve vise les collocations fréquentes verbe-nom dans le français langue seconde écrit, relevées à l'aide de la base de données Les Voisins de *Le Monde*. Les auteures présentent trois différentes études pilotes réalisées à la fois auprès de locuteurs natifs et d'autres locuteurs du français (dont les langues maternelles diffèrent), pour un total de 152 participants. Le test est validé grâce à une épreuve de maîtrise générale et une épreuve de fiabilité. La dernière épreuve, menée auprès de 47 locuteurs du français de langue maternelle différente, consiste dans un test comportant au total de 30 items dont les résultats présentent d'importants écarts pour les participants de niveau B2 et les participants de niveau C1. En plus d'offrir un outil efficace d'évaluation linguistique, cette étude confirme la place essentielle des connaissances productives liées aux collocations, aux niveaux élevés de maîtrise de la langue seconde.

Mots clés : CECR, collocations, connaissances productives, épreuve, français langue seconde

Abstract: The present study investigates the relationship between productive collocation knowledge and advanced levels on the Common European Framework of Reference for Languages (CEFR) scale. More precisely, the potential progression between the B2 and C1 levels of the CEFR is studied through the development of a productive collocation test in second language (L2) French. The test targets frequent verb–noun collocations in written L2 French, identified using the database Les Voisins de *Le Monde*. Three different pilot studies are presented, with both native speakers (NSs) and non-native speakers (NNSs) of French (with different first languages), for a total of 152 participants. The test is validated through a general proficiency test and through a test of reliability. The final testing session,

© *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 74, 4, (November/novembre), 627–649 doi:10.3138/cmlr.2017-0093

carried out with 47 NNSs of French, yielded a test with a total of 30 items, which showed significantly different results for the B2 and the C1 levels. In addition to constituting a useful tool in language assessment, this study confirms the key place of productive collocation knowledge at high levels of L2 proficiency.

Keywords: CEFR, collocations, L2 French, productive knowledge, testing

Applied linguists, second language acquisition (SLA) researchers, and test constructors now agree that knowledge of formulaic language is a highly efficient measure of second language (L2) proficiency, especially at advanced levels. Norbert Schmitt (2012, p. 7) goes as far as to say,

> [i]t seems that mastery of formulaic language takes a long time to acquire and is a hallmark of the highest stages of language mastery. Language testers have picked up on this and often include items which focus on phraseology in their highest-level examinations. Formulaic language is an important element of language overall, perhaps *the* essential element. [emphasis in original]

The term "formulaic language" covers a plethora of phenomena, such as idiomatic expressions, collocations, social routines, and sayings. These linguistic structures have in common that they are multi-word units with a conventional form–meaning mapping, but their definitions and identification criteria vary. The focus of this article is collocations, one of the most frequently occurring categories of formulaic language (cf., e.g., Cowie, 1992; Kjellmer 1987). Collocations are combinations of words that occur frequently together in a given language such as, for example, *keep a secret* (for a definition, see below). In this study, the terms "formulaic language" and "formulaic competence" are used for speaking about the phenomenon in general, whereas the term "collocations" is used in regard to this specific category, which are also the specific linguistic items targeted in this study.

In a paper reviewing research on size and depth in L2 vocabulary research, Schmitt (2014) concludes that the most challenging vocabulary knowledge for L2 learners seems to be *productive collocation knowledge* (one component of vocabulary depth). This conclusion converges with results from the field of SLA (e.g., Durrant & Schmitt, 2009; Nesselhauf, 2003), including Erman, Denke, Fant, and Forsberg Lundell (2015), who show that, in various spontaneous production tasks, collocations are the most difficult category of formulaic language to produce.

As mentioned above, mastery of formulaic language is often assumed to be an efficient indicator of language proficiency with respect to language assessment. The central role of formulaic language is, for

instance, acknowledged in the Common European Framework of Reference for Languages (CEFR). For lexical competence, "fixed frames, phrasal verbs and fixed collocations (e.g. *make a mistake*)" are mentioned among the most important features (Council of Europe, 2001, p. 111). However, the ability to use formulaic language in a (near) native or target-like manner, according to CEFR descriptors, is relegated largely to the very highest levels of proficiency (i.e., C1 and C2), with the exception of unanalyzed chunks, which are expected to be used at the early stages of acquisition. Therefore, we find numerous references to the use and comprehension of formulaic language starting from the C1 level, including "a good command of idiomatic expressions and colloquialisms" (p. 112). This suggests that such mastery characterizes only the highest levels of proficiency and says nothing about the potential development of formulaic competence prior to the C1 level.

Although the CEFR has had wide-ranging influence on language teaching and assessment, especially in a European context, Hulstijn (2007) points out that the levels described in the CEFR are to be considered with caution. This is because the empirical foundation of the CEFR is not built on evidence from developmental studies but, rather, relies on experienced teachers' assessments and practice (i.e., teachers' perceptions of learners' capabilities at different levels). In other words, many experienced teachers and language educators would probably agree on the fact that idiomatic features of language are characteristic of the C1 level; however, this is not based on empirical research, comparing B-level with C-level learner performance. Forsberg and Bartning (2010) conducted a pilot study on written L2 French to investigate the relationship between CEFR levels and linguistic features. Among other things, it was found that formulaic language significantly separated groups at the most advanced levels, while morphosyntactic criteria failed to distinguish groups after the B1 level. These findings suggest that it is worth pursuing the relationship between advanced CEFR levels and formulaic language. The present study thus contributes to filling this knowledge gap, by relating the B2 and C1 levels of the CEFR to results on a productive collocation test. Moreover, examining these two levels is interesting given that the CEFR insists on formulaic competence starting from the C1 level, leaving open the question of the extent to which this competence develops at previous levels of proficiency. In order to address this knowledge gap, we have set about creating a test of productive collocation knowledge for L2 French.

Within applied linguistics, attempts have been made to create collocation tests, although none has been linked directly to the CEFR levels. Looking first at L2 English, which has arguably received the most

attention, there have been attempts to develop both receptive (Eyckmans, 2009; Gyllstad, 2007) and productive (González Fernandez & Schmitt, 2015; Mizrahi & Laufer, 2010) measures of collocation knowledge. There is still, however, a lack of tests in languages other than English. In L2 French, a first attempt was made by Forsberg Lundell and Lindqvist (2014) to create a productive collocation test. In the current article, we aim to build on this research with a new productive test of collocation knowledge. We have opted to target productive–and not receptive–knowledge largely because we are interested in investigating collocation knowledge at the highest levels of proficiency. With such highly proficient speakers, the use of receptive tests may very well result in ceiling effects (see Schmitt, 2010). Through the development of a test of productive collocation knowledge, we seek the answer to this research question: Is there a progression in productive collocation knowledge between the B2 level and C1 level of the CEFR?

Besides responding to a theoretical interest in the linguistic correlates of the CEFR, the study also aims to provide a test that could be useful in language assessment, not least with respect to language entrance exams prior to university, which often place students at a given CEFR level, either via self-assessment or via a holistic judgment made by experienced evaluators. The test could also potentially be used for training purposes or as an assessment tool at advanced levels of French study at university.

Previous research

Formulaic language has been investigated in L2 learning from a wide variety of perspectives: Researchers have examined which formulaic units are used by L2 speakers in writing and in speech (including in different genres, with particular attention having been paid to academic discourse), they have administered off-line measures in order to gauge collocation knowledge, and they have sought to determine how L2 speakers process formulaic units in their L2 (using eye tracking and self-paced reading tasks). In what follows, we will provide an overview of only the research that is relevant for the development of a task of productive L2 collocation knowledge.

Defining collocations

There are various definitions of collocation, although most researchers would agree that collocations consist of words that commonly appear together but can often be perceived as arbitrary word combinations from the L2 users' perspective, such as *make* a deal (and not *do) and *have* breakfast (and not *eat). There are two main ways of defining a

collocation: the phraseological definition and the statistical definition. The phraseological definition identifies sequences based on their syntactic, semantic, and pragmatic restrictions. For instance, Erman and Warren (2000) use the criterion "restricted exchangeability" to determine the fixed status of a sequence. Restricted exchangeability implies that one of the components of the sequence cannot be changed for a synonym without loss of idiomaticity or meaning. An example would be that one says "choose words" and not "*select words." Studies that have used phraseological methods include Howarth (1998), Nesselhauf (2003), and Bolly (2008).

The statistical definition, which will be used in this paper, considers a collocation to be any word combination in which the included words appear together more often than by chance (Paquot & Granger, 2012). Within the statistical approach it is common to use a mutual information (MI) score to define a collocation. This measure provides a numerical indicator of the cohesiveness of any given word combination within a corpus, that is, the strength of association between two words in a specific corpus. It is also common to use frequency of occurrence of the word combination in a corpus as part of the definition, with thresholds for frequency of occurrence (cf. Ellis, Simpson-Vlach, & Maynard, 2008). This is to measure how well-established a certain word combination is, within a certain corpus or community of practice. Both of these measures will be used in this study.

Previous research on productive collocation knowledge and use in L2

In this section, we review research on collocations in productive tests or written production. Previous research on collocation knowledge and use has shown that it is a late feature in L2 learning, which often seems to distinguish native speakers (NSs) from non-native speakers (NNSs), in terms of both the number of collocations used and the quality (i.e., well-formedness) of those collocations (e.g., Arnaud & Savignon, 1997; Forsberg Lundell & Lindqvist, 2012; Lewis, 2008; Mizrahi & Laufer, 2010, 2011). Mizrahi and Laufer (2011), for example, analyzed lexical use in free writing and found that highly advanced learners of L2 English achieved native-like levels by other lexical measures, namely lexical richness (as measured by the Lexical Frequency Profile [Laufer & Nation, 1995]) and lexical variation (as measured by the type/token ratio), but not by the use of collocations. Similarly, Forsberg Lundell and Lindqvist's (2012) study showed that advanced learners of L2 French reached native-like levels on lexical richness, but they did not produce formulaic sequences (the term used in their study) within the native-speaker range (the range of results produced by a native-speaker

control group). Furthermore, in Forsberg Lundell and Lindqvist (2014), the results also pointed to differences between NSs and NNSs regarding collocation knowledge. However, there were several NNSs who performed within the NS range.

Durrant and Schmitt (2009) examined the use of highly frequent English collocations by NSs and NNSs, studying written texts. Collocations were measured with respect to both MI scores and frequency of occurrence. The authors focused on noun–noun and adjective–noun combinations. The results showed that the NS writers used more low-frequency combinations than did NNS writers. Furthermore, it was demonstrated that the NNS writers used fewer collocations with high MI-scores than the NS writers. Durrant and Schmitt (p. 175) concluded that:

> [a]dvanced non-native phraseology differs from that of natives not because it avoids formulaic language altogether but because it overuses high-frequency collocations and underuses the lower-frequency, but strongly-associated, pairs characterized by high mutual information scores. Since the latter sort appear (intuitively, and on the psycholinguistic evidence presented by Ellis et al.) to be highly salient for native speakers, their absence may be what creates the feeling that non-native writing lacks "idiomaticity."

While learners can be sensitive to both frequency of occurrence and the cohesiveness of a sequence, it appears from the study above that it is not primarily the cohesiveness of a word combination that will affect how easily it is produced by L2 learners, but rather its frequency of occurrence.

Research on linguistic correlates and the CEFR

In this section, we review work that has, as we have, tried to relate linguistic data to the CEFR levels. Since the publication of the CEFR in 2001, numerous attempts have been made to identify language-specific linguistic correlates for each of the six levels of the framework. One result of this work has been the development of reference-level descriptors, which "identify the specific forms of any given language (words, grammar, etc.), the mastery of which would correspond to the competences defined by the CEFR at each of the six reference levels" (Saville & Hawkey, 2010, p. 3). Within the English Profile Programme at the University of Cambridge (Saville & Hawkey), extensive work has been carried out in order to identify criterial features, defined as "certain linguistic properties that are characteristic and indicative of L2 proficiency at each level, on the basis of which examiners make their practical

assessments" (Hawkins & Buttery, 2010, p. 3). Included among these linguistic properties was vocabulary, with researchers initially working to compile what were termed "English Profile Wordlists." Interestingly, the term "Wordlists" was ultimately abandoned, in favour of the title "English Vocabulary Profile." According to Capel (2012), this change in terminology reflects the greater presence of multi-word units in the vocabulary profiles, particularly at the highest levels of L2 proficiency (C1 and C2). This observation is also reflected in the CEFR itself, where mentions of mastery of idiomatic expressions (Council of Europe, 2001, p. 111), colloquialisms (p. 111), and appropriate formulations (p. 110) are prevalent at levels C1 and C2.

Two recent studies have set out to examine the relationship between collocation knowledge and overall proficiency as measured by either the CEFR or the TOEFL. In a large-scale study, Hulstijn, Schoonen, de Jong, Steinel, and Florijn (2012) examined how a range of linguistic competences – including collocation knowledge – succeeded in discriminating between learners of Dutch independently judged to be at a B1 or B2 level in oral production. Of particular interest for the current study is the finding that the measure of vocabulary knowledge (involving both knowledge of individual words and knowledge of collocations) was consistently the best predictor of whether an individual was judged to be at a B1 or a B2 level for a speaking task (Hulstijn et al., p. 214). Nizonkiza (2012) investigated controlled productive knowledge of verb–noun collocations in L2 English, with special attention paid to the frequency of the base of the collocation and the proficiency level of the L2 speaker. Participants were students majoring in English at a university in either Belgium or Burundi who were required to complete both the TOEFL and a collocation test. The format of Nizonkiza's collocation test resembles our own, and it required participants to fill in the missing collocate (i.e., the verb) in the context of a sentence; for each item, the first two letters of the verb were provided. Results revealed that when participants were grouped into five levels according to their overall L2 proficiency, results from the collocation test were also significantly different, with the exception of the first two proficiency levels. Nizonkiza argues that this suggests a significant and positive relationship between collocational knowledge and overall proficiency.

In short, research in both SLA and language assessment stresses the important role of collocation knowledge, and promising attempts have been made to relate this type of knowledge to the CEFR. In what follows, we describe our study, which aims at creating a productive collocation test for L2 French, related to the CEFR. We then use the results from the creation of this test in order to examine the potential

progression in collocation knowledge from the B2 to the C1 level of the CEFR.

Method: Developing the instrument

In this section, we provide details on the development of the collocation test. We begin by describing how items were originally selected, as well as the original test format (see Forsberg Lundell & Lindqvist, 2014, for additional details). We then describe in detail the current study, covering the three pilot tests undertaken as well as changes made to the test as a result of each pilot study. The section ends with a presentation of the testing of the final version of the collocation test. It should be noted that in the first phases of the test construction, the only participants were Swedish learners of French, but in the later phases the test was used with participants of many different first languages (L1s) and its final version is intended to be used irrespective of L1.

Item selection and test format

As mentioned above, a first version of this French verb–noun collocation test was created by Forsberg Lundell and Lindqvist (2014). As no similar test existed in French, they followed the methodology used in previous research in which L2 English collocation tests had been designed. Specifically, when selecting which items to include in the test, Forsberg Lundell and Lindqvist (2014) followed Gyllstad's (2007) procedure (see below). With respect to format, their test was modelled after the measure created by Mizrahi and Laufer (2010), who had also designed a productive test; Gyllstad, on the other hand, had designed a receptive test. Various steps were undertaken when creating the collocation test. The first step was to decide on the nature of the collocations to include. In line with Gyllstad, the authors opted for verb–noun collocations, such as *rendre justice* ("to do justice"). This choice also coincides with that of many other studies on SLA and collocations (e.g., Boers, Demecheleer, Coxhead, & Webb, 2014; Laufer & Waldman, 2011; Nesselhauf, 2003), which points to their relevance and difficulty in L2 learning. Moreover, the authors decided to base the selection of items on written French data, since the test was to be taken in the written mode. The procedure was as follows:

1. One hundred and fifty frequent nouns were randomly selected from the first three frequency bands, containing the 3,000 most frequently used French words in written French, available on the web site "Frequency lists for teachers/researchers" (https://www.lextutor.ca/freq/lists_download/). The frequency bands are based on a 50-million-word database of two French-language newspapers: *Le Soir* (Belgium) and *Le Monde* (France).

2. The database Les Voisins de *Le Monde* (http://redac.univ-tlse2.fr/voisinsdelemonde/) was used to search for collocating verbs for the 150 nouns.[1] Les Voisins de *Le Monde* contains 200 million words and is based on all the issues of the newspaper *Le Monde* from 1991–2000. Importantly for the purposes of Forsberg Lundell and Lindqvist's (2014) study, the database provides both frequencies and MI scores.
3. Two selection criteria were applied in the search for relevant test items: the MI-score threshold was set at three, in accordance with many other studies in the field (e.g., Ellis et al., 2008; Siyanova & Schmitt, 2008); and the frequency threshold was set at 200, which, given the size of the database, yields a minimum frequency of one per one million words.
4. Since the test was supposed to target the most advanced levels of L2 proficiency, it was decided that items that appeared to be relatively easy would be eliminated. It is perhaps not surprising that L2 collocation research has shown that collocations that are equivalent in the L1 and the L2 convey acquisition advantages (cf. Leśniewska & Witalisz, 2007; Wolter & Gyllstad, 2011). Therefore, given that the targeted participants for Forsberg Lundell and Lindqvist (2014) were all native Swedish speakers, the researchers removed all collocations that were translation equivalents (or, as some would call them, "cognates") between Swedish and French. Having applied all these criteria, 70 verb-noun combinations were left.
5. In order to construct the test, example sentences for each of the 70 remaining collocations were identified, using www.google.fr. As explained in Forsberg Lundell and Lindqvist (2014), 10 NSs and 10 NNSs of French then completed the test. The items for which less than half of the NSs had provided a collocation were removed. This resulted in 46 items in the final test.
6. This final test was distributed to 15 very advanced L2 users of French, residing in France. The results of this test were presented in Forsberg Lundell and Lindqvist (2014). The test took the form of a fill-in-the-gap test, following Mizrahi and Laufer (2010), as can be seen in the following examples:

(1) a. *L'ONU est fermement résolue à* ***r____ justice*** *aux victimes du génocide*. (The United Nations are determined to d___ justice to the victims of the genocide.)

b. *Si les petites entreprises* ***é_____ des difficultés*** *à embaucher, c'est aussi parce qu'elles disposent de moyens qui sont moindres*.

(If small companies h___problems in hiring, it's also because they have smaller budgets.)

c. *L'industrie porno entend* ***m*****_____*****un terme*** *au piratage d'ici 2012*. (The porn industry is determined to p____an end to pirate copying by 2012.)

Whereas Mizrahi and Laufer (2010) provided the two first letters for each blank, Forsberg Lundell and Lindqvist (2014) decided that only the first letter of each response would be presented in order to avoid ceiling effects.

However, as discussed below, this format was changed slightly in the context of the current study, the aim of which was to create a validated test of productive collocation knowledge, related to the B2 and C1 levels of the CEFR. In the rest of this section, we detail in a linear fashion the three pilot tests that led to the development of the final test presented in this study. The next section will focus on the presentation of the results from the testing of the final version.

Developing and piloting a validated test

Having decided that we wanted to further develop the Forsberg Lundell and Lindqvist (2014) test into a validated test, a procedure with different pilots and different populations was undertaken.

Pilot 1

The first pilot study was conducted with university teachers of French in Sweden. Seventeen participants, including both NSs and NNSs, took the test. Although the Forsberg Lundell and Lindqvist (2014) test was the starting point for this project, we did not administer the 46-item test that they used. Instead, in order to apply new selection criteria for item identification, we returned to the original item pool reported in Forsberg Lundell and Lindqvist (2014). The most important difference was that the new test was going to be distributed to speakers of many different L1s, in order to enhance its usefulness, which implies that in the selection of test items, translation equivalents between French and Swedish would not be eliminated. Thus, 88 items from the original sample collected by Forsberg Lundell and Lindqvist (2014) were included in Pilot 1. The test instructions, which were revised in later versions of the test (see below), were (as translated from Swedish):

This test contains 88 sentences. In each sentence, one verb is missing. Your task is to fill in the verb that you think is most appropriate in the context. No help is permitted! The first letter of the verb is provided. Good luck!

In coding the answers provided, each expected response was awarded one point. For all remaining responses, the new (i.e., unexpected) verb-noun combinations were verified in Les Voisins de *Le Monde*, and their frequency and MI scores were recorded. Using this information, the following scoring method was applied, dividing answers into two main categories: 1. collocations and 2. non-collocations, with respective subcategories.

1. Collocation

   a. The most expected collocation, i.e. the strongest collocation: 1 point (an MI score > 3 and a frequency score > 200)
   b. Not the most expected collocation, but a collocation (an MI score > 3 and a frequency score > 5[2]) according to searches in Les Voisins de *Le Monde*. For example, in the case of the item "Il faut savoir ***porter*** un regard bienveillant sur autrui sans avoir de préjugés [You need to be able to **view** others in a benevolent and non-judgmental manner]," it was found that the verb *poser* ("to lay") also met the MI score (> 3), typically used as an MI threshold for collocational status, although the frequency of occurrence was much lower than the frequency for the expected collocate, porter: 1 point

2. Non-collocation

   a. Any verb–noun combination with an MI score below 3 was considered as a free combination and thus not a collocation: 0 points
   b. Verb–noun combinations not found in Les Voisins de *le Monde*: 0 points

As explained below, these coding criteria were modified in later versions.

The next step in Pilot 1 was to determine which test items to keep for further testing. It was decided that in cases where all participants provided a given collocation (i.e., following the criteria for category 1 above), the item would be kept. Items resulting in non-collocation answers were discarded.

Pilot 2

As a result of the first pilot test, 81 items were retained. In order to test this new version, we wanted to recruit a larger number of NSs and NNSs, so we decided to recruit these participants through a university in France. Indeed, given that the ultimate target population for the test is high-level learners of L2 French from a variety of L1 backgrounds, we looked to conduct this second pilot test in precisely such a context and with a similar participant pool.[3] To that end, 36 NSs and 41 NNSs of French living in the same community in France completed the test. The 36 NSs (five men and 31 women) were all enrolled in either undergraduate or graduate

didactics courses at a university, and completed the task during a class period. They were given 45 minutes to do so. These speakers were on average 23.5 years old (*SD* = 3.6). For the NNSs, we sought to identify individuals having attained the C level of the CEFR in French. Because C is an advanced level, finding a sufficient number of such participants can be challenging. For this reason, we adopted two approaches to recruiting such NNSs. First, two intact classes of students enrolled at the French language institute at the same university in France and who had been placed, on the basis of the entrance exam, at the C level, completed the task (*n* = 24). These participants were an average of 23.1 years old (*SD* = 3) and had spent an average of 13.1 months (*SD* = 11) in France at the time of testing. Seven of the participants were men, whereas 17 were women, and they reported nine different L1s. Second, NNSs living in the same community were contacted through their links to the local university. Twenty-two such speakers agreed to complete the task. In order to verify their level in French, we asked these speakers to self-evaluate their global proficiency. Five of the 22 estimated their level to be below the C level; we therefore excluded these speakers from the final analysis. The 17 remaining speakers had five different L1s and included five men and 12 women. On average, they were 35.4 years of age (*SD* = 9.7) and had been living in France for 9.9 years (*SD* = 7.7). Taken together, 41 NNSs completed the test.

Responses were scored in the same way as for Pilot 1 (see categories 1–2), and results showed that the NSs answered on average 60 items (of the 81 possible) with collocations (*SD* = 10.11), whereas the NNSs responded to only an average of 36 items with collocations, with a standard deviation two times higher than was found for the NSs (20.37). These results raised several questions. First, for certain items, a large number of responses were counted as collocations, following the rather generous scoring procedure adopted for the first pilot test. For example, this was the case with the target collocation *atténuer l'effet de serre*, which was tested using the carrier phrase provided in (2).

(2) *Les forêts peuvent a*_____ *l'effet de serre*.

Twelve verbs other than *atténuer* were provided as responses to this item, of which nine were scored as collocations following the scoring procedure outlined for Pilot 1. Despite having benefited from generous scoring, the NSs in this pilot were far below ceiling levels (scores ranged from 31 to 76 out of 81 possible points on the full test). This is a clear difference from Pilot 1, which may reflect the higher level of education of most of the NSs tested in Sweden. Regardless of the explanation, this wide range suggested to us that the current version of the test was too difficult, even for NSs enrolled in university courses.

As a result of these observations, the test was modified in two ways. First, we removed all items for which fewer than half of the NSs had provided a collocation. This resulted in the removal of 40 items, leaving us with 41. Second, due to the apparent difficulty of the test, we decided to test two answer formats: one version in which the first letter of each verb would be provided (as was the case in the current format) and one version in which the first two letters would be provided (note that this two-letter format was also adopted by Mizrahi & Laufer [2010]). In addition to these two changes to the test itself, we made the decision to change our scoring procedure. In scoring results from this pilot test, it became clear to us that the approach adopted for Pilot 1 allowed for responses that were not clearly collocations. Given that the goal of the current test was to evaluate collocation competence at high levels of L2 French competence, we decided that in the next phase of testing, 1 point would be awarded only for verb–noun combinations with an MI score of 3 or above and occurring at least one time per 1 million words (> 200 occurrences in Les Voisins de *Le Monde*), counting only these as collocations. This differs from Pilot 1, where combinations with an MI score of at least 3 and a frequency of 5 or greater were also counted as collocations. This new, stricter scoring procedure allowed for only one possible collocation for 38 of the 41 remaining items and two possible collocations for the remaining three items.

Pilot 3

The principal aim of the third pilot study was to determine whether the new 41-item test could successfully tap NSs' collocation knowledge. In other words, before using the test with high-level NNSs, we wanted to ascertain whether NSs would score well on this new version. To this end, we recruited a total of 58 NSs of French from the same university in France where Pilot 2 was conducted; none of the participants had taken part in previous piloting. All speakers were enrolled in a Masters degree in law, humanities, business management, or mathematics. Twenty-one participants were men, 36 were women, and one did not provide their gender. These speakers were on average 23 years old (SD = 1.6) and were assigned at random to either the one-letter or the two-letter condition. Participants were allowed up to one hour to complete the test, although most had finished after 30 minutes. Participants were paid €10 for their time.

Responses were scored using the strict scoring procedure mentioned above. Looking first at the results for the version of the test in which only one letter of the expected word was provided, we see that the 29 participants answered on average 32 (of 41 = 78%) questions with

collocations, with a range of 23 to 38 total collocations. By comparison, the 29 participants who had the two-letter format provided collocations to an average of 36 (= 87.8%) items, with a range of between 29 and 40 correct responses per individual. The proportion of collocations was therefore higher for both versions of the test than in Pilot 2, and results from the version with the two-letter answer format were higher compared to those from the one-letter version. After having examined the overall proportion of collocations provided, we then looked at how the participants performed on individual items, comparing the two versions of the test. Looking first at the version with one letter provided for each response, 17 items were answered with collocations by at least 90% of the 29 participants (i.e., 26 of the 29 participants). This number climbed to 31 on the two-letter version. As a result, for the final version of the test, we opted for the two-letter format and retained only the 31 items for which 90% of the participants in Pilot 3 provided a collocation.[4]

Testing the final version: Participants and procedures

The final version of the collocation test contains 31 items which, as seen in the results from Pilot 3, resulted in high rates of collocations from a group of NSs in a university context. Having established that NSs perform well on these items, we then turned to high-level learners of French for the final testing phase in order to determine whether the final version of the collocation test would show a progression between the B2 and the C1 level in L2 learners of French. A total of 50 NNSs enrolled at one of three universities located in the same geographic region of France participated in this final testing phase. Learners were contacted via the French language institutes at the different universities, and most participants intended to pursue their studies in France. Participants included 41 women and nine men with 17 different L1s. Learners with a range of different L1s were included, as we wanted the test to be useful to many different L1 learner populations. These learners were on average 25.6 years old (SD = 4.5) and had been living in France for an average of 21.8 months (SD = 15). Participants were offered a gift certificate for €10 to compensate them for their time.

In order to determine whether the collocation test would show a progression between levels B2 and C1, each participant had to complete a proficiency test against which the results from our collocation test could be compared. This was done with the intention of measuring the test's concurrent validity (cf. Bachman, 1990). This type of validation is obtained through correlating results from two tests supposed to measure second language proficiency, but not the same construct.

For this, we chose the proficiency test for level C proposed by the Folkuniversitetet.[5] This test contains 40 questions that target predominantly grammatical and lexical competence, with results dividing participants into five levels: B1+, B2, B2+, C1, and C1+; participants obtaining fewer than nine total points are reoriented toward the test for the B level. This test has been validated by the Folkuniversitetet and was chosen for the current project mainly because it had been used in earlier studies where the CEFR level needed to be controlled for. (cf. Falk, Lindqvist, & Bardel, 2015). In (3), two example sentences from the test are provided:

(3) **10.** Mon chef n'acceptera jamais que je..... autant de vacances!
A prenne
B prendrai
C prends

11. Dépêchons-nous de rentrer avant qu'il..... à pleuvoir.
A se mettra
B se mette
C se met

We wanted to use a test that would be relatively quick to administer, as participants would have to complete both the general CEFR proficiency task and the collocation test. The two tasks together required no more than 45 minutes per participant, and the order of the tasks was counterbalanced. According to the results from the Folkuniversitetet test, three participants were at the B1 level, 26 were at the B2 level, and 21 were at the C1 level. The three participants having obtained B1-level scores were removed from the data set, leaving us with a total of 47 participants.

Results from the final test

The participants were divided into two groups, based on their scores on the CEFR test designed by the Folkuniversitetet (FU level) (see Table 1). There were 26 participants in the B2 group and 21 participants in the C1 group. The mean collocation score was 19.7 in the B2 group and 25.7

Table 1: Collocation score divided by CEFR-level (N = 47)

FU Level	*M (SD)*	*SE*	Min–Max
B2 (n = 26)	19.7 (4.7)	0.9181	11–30
C1 (n = 21)	25.7 (4.1)	0.9054	17–31

in the C1 group. The difference observed between the two groups on the collocation test, based on a *t*-test, was highly significant: $p < 0.0001$, $t = -4.64$. This suggests that productive collocation knowledge, as measured by this test, shows a significant progression between B2 and C1 levels.

A Pearson correlational analysis (Figure 1) also yielded a significant, moderate correlation between the two variables of collocation score and CEFR-level (FU score) ($r = 0.647$, $p = < 0.001$), which offers supplementary evidence of the association between CEFR level and collocation knowledge, as measured by the Folkuniversitetet general proficiency test and the productive collocation measure reported on in this article.

With respect to the test's reliability, we also conducted a Cronbach's alpha analysis. According to Schmitt (2010, p. 185), internal consistency is a frequently used measure of reliability, especially due to practical reasons, which makes test–retest procedures quite difficult to conduct. The Cronbach's alpha analysis showed that the test demonstrates a good internal consistency ($\alpha = 0.844$).[6] One of the items (item 10) was discarded by the Cronbach's alpha analysis because it displayed no

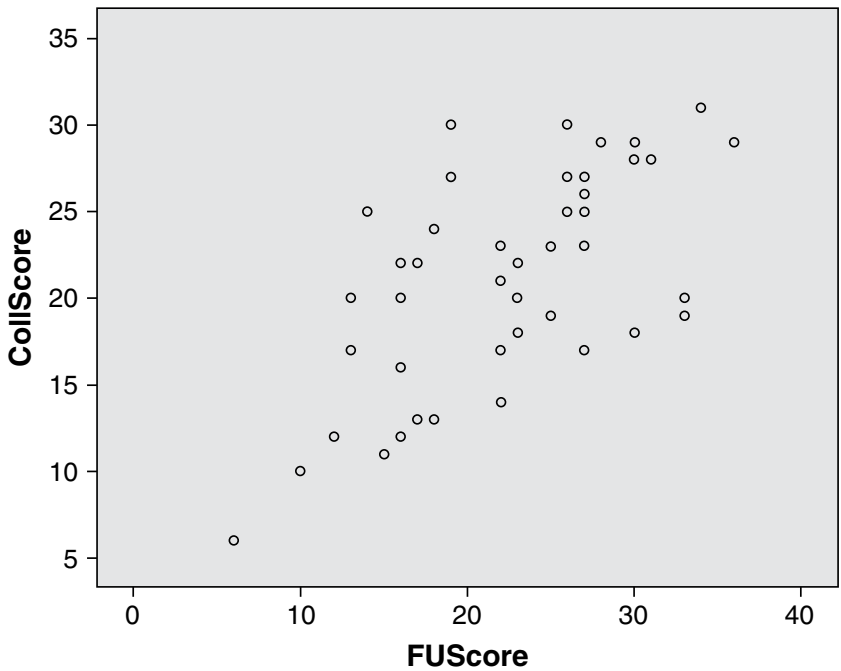

Figure 1: Scatterplot of correlation between collocation score and CEFR-level (FU score)

variance whatsoever (all the participants provided collocations for this item). The item in question was *garder un secret* ("keep a secret"). This item was consequently removed from the test, leaving a total of 30 items in the final test format.

As Schmitt (2010, p. 165) justly acknowledges, there is no set answer for the tricky question of how many items should be included in a vocabulary test. The researcher has to combine practical constraints with reliability and validity constraints. The test that we have described is indeed practical, as it requires little time for administration and is easy to score. There is also evidence for its reliability, in the form of a rather high internal consistency (0.844) with a relatively low number of items. Indeed, more items generally result in higher internal consistency. Furthermore, Nation (2001, p. 45) mentions "around 30 items" as a minimum for a reliable test. We purposefully started off with a large pool of items (88 items), knowing that some of them would have to be discarded during the various pilot studies with both NSs and NNSs. Finally, in terms of validity, we have provided initial evidence of concurrent validity by examining the relationship between the productive collocation test and the FU test. Additional measures of validity (by examining, for example, the relationship between performance on the collocation test and on other measures of general proficiency) should be undertaken in the future.

Discussion and conclusion

As may be seen from this study, the process of developing a collocation test for advanced L2 French has been a long one. The most challenging issue for us has been to find informants with a sufficiently advanced level of French, which explains the somewhat modest number of participants.

The most interesting finding of the paper is most certainly that productive collocation knowledge is significantly different in the B2 and C1 level groups, with significantly better scores being seen in the C1-level group. We are thus able to offer an affirmative response to this paper's research question, which asked whether or not there is a progression in productive collocation knowledge between the B2 level and C1 level of the CEFR. This result concurs with the descriptors in the CEFR, in that mastery of idiomaticity is considered characteristic of the C1 level. It should, however, be noted that mastery of collocations is not an all-or-nothing affair with respect to progression from B2 and C1 level. C1 participants are likely to have a higher productive knowledge of collocations (M = 25.7/30) than B2 participants, but B2 participants also display extensive knowledge of collocations (M = 19.7/30). This suggests that

mastery of formulaic knowledge is also characteristic of the B2 level, an observation that is not evident from the CEFR descriptors.

Interesting observations made during the process of creating this test include the variability found among both L1 speakers and L2 speakers with respect to mastery of collocations. Although not addressed in the current project, this variability may very well reflect individual differences in language learning and language use. Research into the impact of individual differences on collocation competence constitutes a promising research area (see Forsberg Lundell & Sandgren, 2013; Granena & Long, 2013).

Another observation concerns the identification of items for the test. Indeed, it appears that the use of statistical identification methods, as was the case in this study, can result in a relatively large number of potential collocates (see discussion with respect to example 3 above). As a result, many items that were initially identified as potential items had to be discarded, since they offered too many alternatives. For scholars working within the field of phraseology, this is not surprising in view of the difficulties surrounding the identification of formulaic units (cf. Forsberg 2010; Myles & Cordier, 2017; Wray, 2002). This gives us interesting indications for further research into the nature of collocations with respect to cohesiveness, restrictions in exchangeability, and optionality. There is still more to add to the discussion on how formulaic language can be defined.

The present study has investigated the relationship between a test of productive collocation knowledge and the B2 and C1 levels of the CEFR, but the data point to further research avenues to be explored. In the future, it would be interesting to conduct an item-by-item analysis, in order to see whether certain items seem to be mastered earlier than others. Is there a hierarchy in terms of collocation difficulty? If yes, can such a hierarchy be explained by factors such as frequency or transparency? This would contribute to the state of the art with respect to L2 acquisition of collocations.

There are, of course, a number of limitations to this research design. Since the first version from Forsberg Lundell and Lindqvist (2014), we have undertaken quite a few steps in the process of developing a valid test of productive collocation knowledge in L2 French. Further steps to improve the test would be to proceed to factor analysis and principal component analysis, in order to investigate the discriminatory power of the test. Another important improvement would be to implement additional proficiency tests for the assessment of the link between collocation knowledge and CEFR proficiency; we acknowledge the limitations of having used only the test from the Folkuniversitetet in this study.

Finally, we would like to emphasize that the present study, apart from offering a new test for use in the assessment of advanced L2 French, also confirms the special place of collocations in very advanced L2 acquisition and their role in describing (non) native-like language use in an L2. These results are very much in line with those suggested by Hulstijn et al. (2012) and Nizonkiza (2012) and also confirm the initial statement by Schmitt (2012), in which he reserves such a key place for formulaic language within language testing and applied linguistics more generally.

Correspondence should be addressed to **Fanny Forsberg Lundell**, Department of Romance Studies and Classics, Stockholm University, Stockholm, Sweden, SE-10691; e-mail: fanny.forsberg.lundell@su.se.

Notes

1 Since the test was to be taken in a written mode, it was deemed adequate to use written reference corpora. It was also important to limit the corpus to Metropolitan French, since the test is developed with native speakers of Metropolitan French. Les Voisins de *Le Monde* is the largest existing written corpus of Metropolitan French and, in addition, it has the advantage that it provides MI score figures.

2 This figure is much lower than >200, which was set as a frequency criterion in Forsberg Lundell & Lindqvist (2014). However, we discovered that quite a few word combinations produced by the participants were of this kind: word combinations that were collocations with respect to MI score (>3) but that were not frequent. At this point, we thought that it was difficult to discard them as "non-collocations," in view of their MI score, their use by native participants, and their presence in the reference corpus.

3 It should be acknowledged that some of these speakers will have L1s that are typologically closer to French than others, containing more collocational cognates. Collocational cognates have proven to facilitate the processing and acquisition of collocations in L2 (Leśniewska & Witalisz, 2007; Wolter & Gyllstad, 2011). For speakers of closely related languages (e.g., Spanish and Italian), the test may thus be easier. Since we wanted to construct a test that could be used in many contexts, the cross-linguistic bias constitutes a limitation.

4 Missing answers were given a score of zero.

5 This test is a written placement test, focusing mainly on grammar and lexicon (from the Open University in Sweden). It is part of a set of tests developed in accordance with the CEFR scale A1–C2 (Council of Europe, 2001) and available in English, French, Italian, Spanish, and German.

6 In addition to calculating the traditional Cronbach's alpha, we also calculated summability for the test, a new measure for internal consistency, proposed by

Goeman & De Jong (2018). The summability score obtained was 0.3, which, according to the authors, can be considered to be moderate summability. However, this measure has not yet been used sufficiently to establish thresholds for weak, moderate, and high summability. The test would benefit from closer examination of the individual items in order to strengthen its validity.

References

Arnaud, P.J.L., & Savignon, S.J. (1997). Rare words, complex lexical units and the advanced learner. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 157–173). Cambridge, England: Cambridge University Press.

Bachman, L.F. (1990.) Fundamental considerations in language testing. Oxford, England: Oxford University Press.

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

Bolly, C. (2008). *Les unites phraséologiques – un phénomène linguistique complexe ? Séquences (semi-) figées construites avec les verbes prendre et donner en français écrit L1 et L2 Approche descriptive et acquisitionnelle* (Doctoral dissertation. Université Catholique de Louvain). Retrieved from https://dial.uclouvain.be/pr/boreal/object/boreal:19625/datastream/PDF_08/view

Capel, A. (2012). Completing the English Vocabulary Profile: C1 and C2 vocabulary. *English Profile Journal*, *3*, 1–14. https://doi.org/10.1017/s2041536212000013

Council of Europe. (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment. Strasbourg, France: Language Policy Unit. Retrieved from https://rm.coe.int/16802fc1bf

Cowie, A.P. (1992). Multiword lexical units and communicative language teaching. In P. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 1–12). London, England: Palgrave MacMillan.

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching*, *47*(2), 157–177. https://doi.org/10.1515/iral.2009.007

Ellis, N., Simpson-Vlach, C.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

Erman, B., Denke, A., Fant, L., & Forsberg Lundell, F. (2015). Nativelike expression in the speech of long-residency L2 users: A study of multiword structures in the speech of L2 English, French and Spanish. *International Journal of Applied Linguistics*, *25*(2), 160–182. https://doi.org/10.1111/ijal.12061

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, *20*(1), 29–62. https://doi.org/10.1515/text.1.2000.20.1.29

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

Falk, Y., Lindqvist, C., & Bardel, C. (2015). The role of L1 explicit metalinguistic knowledge in L3 oral production at the initial state. *Bilingualism: Language and Cognition*, *18*(2), 227–235. https://doi.org/10.1017/S1366728913000552

Forsberg, F. (2010). Using conventional sequences in L2 French. *International Review of Applied Linguistics in Language Teaching*, *48*(1), 25–50. https://doi.org/10.1515/iral.2010.002

Forsberg, F., & Bartning, I. (2010.) Can linguistic features discriminate between the CEFR-levels? A pilot study on written L2 French. In I. Bartning, M. Martin, & I. Vedder (Eds.) *Communicative proficiency and linguistic development. Intersections between SLA and language testing research* (pp. 133–158). Eurosla Monograph Series 1.

Forsberg Lundell, F., & Lindqvist, C. (2012). Vocabulary aspects of advanced L2 French: Do lexical formulaic sequences and lexical richness develop at the same rate? *Language, Interactions and Acquisition*, *3*(1), 73–92. https://doi.org/10.1075/lia.3.1.05for

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

Forsberg Lundell, F., & Sandgren, M. (2013). High-level proficiency in late L2 acquisition: Relationships between collocational production, language aptitude and personality. In G. Granena & M. Long (Eds.), *Sensitive periods, aptitudes and ultimate attainment in L2* (pp. 231–258). Amsterdam, Netherlands: Benjamins.

Goeman, J.J., & De Jong, N.H. 2018. How well does the sum score summarize the test? Summability as a measure of internal consistency. *Educational Measurement: Issues and Practice*, *37*(2), 54–63. https://doi.org/10.1111/emip.12181

González Fernandez, 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. https://doi.org/10.1075/itl.166.1.03fer

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

Gyllstad, H. (2007). *Testing English collocations. Developing receptive tests for use with advanced Swedish learners* (Doctoral dissertation, Lund University, Sweden). http://lup.lub.lu.se/search/ws/files/5893676/2172422.pdf

Hawkins, J.A., & Buttery, P. (2010). Criterial features in learner corpora: Theory and illustrations. *English Profile Journal*, *1*, 1–23. https://doi.org/10.1017/s2041536210000103

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics 19*(1), 24–44.

Hulstijn, J.H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *Modern Language Journal*, *91*(4), 663–667. https://doi.org/10.1111/j.1540-4781.2007.00627_5.x

Hulstijn, J.H., Schoonen, R., de Jong, N.H., Steinel, M.P., & Florijn, A. (2012). Linguistic competences of learners of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the Common European Framework of Reference for Languages (CEFR). *Language Testing*, *29*(2), 203–221. https://doi.org/10.1177/0265532211419826

Kjellmer, G. (1987). Aspects of English collocations. In W. Meijs (Ed.), *Corpus linguistics and beyond*. Proceedings of the Seventh International Conference on English Language Research on Computerized Corpora (pp. 133–140). Amsterdam, Netherlands: Rodopi.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*(2), 647–672. https://doi.org/10.1111/j.1467-9922.2010.00621.x

Leśniewska, J., & Witalisz, E. (2007). Cross-linguistic influence and acceptability judgments of L2 and L1 collocations: A study of advanced Polish learners of English. In L. Roberts, A. Gürel, S. Tatar, & L. Marti, *EUROSLA yearbook 7* (pp. 27–48). Amsterdam, Netherlands: John Benjamins.

Lewis, M. (2008). *The idiom principle in L2 English* (Doctoral dissertation, Stockholm University, Sweden).

Mizrahi, E., & Laufer, B. (2010). Lexical competence of highly advanced L2 users: Is their collocation knowledge as good as their productive vocabulary size? Paper presented at Eurosla 20, Reggio Emilia, Italy.

Mizrahi, E., & Laufer, B. (2011). Does one need a native-like productive lexical knowledge to pass for a native writer? Paper presented at Eurosla 21, Stockholm, Sweden.

Myles, F., & Cordier, C. (2017). Formulaic sequence (FS) cannot be an umbrella term in SLA. *Studies in Second Language Acquisition*, *39*(1), 3–28. https://doi.org/10.1017/S027226311600036X

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press.

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

Nizonkiza, D. (2012). Quantifying controlled productive knowledge of collocations across proficiency and word frequency levels. *Studies in Second Language Learning and Teaching*, *2*(1), 67–92. https://doi.org/10.14746/ssllt.2012.2.1.4

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130–149. https://doi.org/10.1017/s0267190512000098

Saville, N., & Hawkey, R. (2010). The English Profile Programme – The first three years. *English Profile Journal*, *1*(1), 1–14. https://doi.org/10.1017/s2041536210000061

Schmitt, N. (2010). *Researching vocabulary. A vocabulary research manual*. Basingstoke, England: Palgrave Macmillan.

Schmitt, N. (2012). Formulaic language and collocation. *The encyclopedia of applied linguistics*. https://doi.org/10.1002/9781405198431.wbeal0433

Schmitt, N. (2014). Size and depth of vocabulary knowledge. What the research shows. *Language Learning*, *64*(4), 913–951. https://doi.org/10.1111/lang.12077

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

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

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