



A Corpus-driven Design of a Test for Assessing the ESL Collocational Competence of University Students

MARÍA MORENO JAÉN*
Universidad de Granada

ABSTRACT

This paper reports an assessment of the collocational competence of students of English Linguistics at the University of Granada. This was carried out to meet a two-fold purpose. On the one hand, we aimed to establish a solid corpus-driven approach based upon a systematic and reliable framework for the evaluation of collocational competence in ESL. On the other hand, it was our intention to determine whether students' collocational thresholds were acceptable. Thus, after revising the theoretical construct of the notion of "collocation", we accomplished the selection of items drawn from data provided by the Bank of English and the British National Corpus, a procedure intended to ensure the scientific quality of the test design propounded. We designed an 80-item test to assess competence in both the receptive and productive collocational aspects of the written skill. Results revealed that students possess a poor collocational competence, the scores concerning the productive items being, as expected, significantly lower than the receptive ones.

KEYWORDS: collocations, testing, corpus-driven, ESL university students.

* *Address for correspondence:* Departamento de Filología Inglesa y Alemana, Facultad de Filosofía y Letras, Campus de Cartuja s/n, 18071 Granada. E-mail: mmjaen@ugr.es.

I. BACKGROUND

For many years the lexical component of the language has been a neglected aspect in the field of Applied Linguistics (Zimmerman, 1997). However, we have witnessed such an increase in the number of studies devoted to this issue in the last 20 years (see, for example, Bogaards & Laufer, 2004; Carter, 1987; Lewis, 1993; McCarthy, 1990; Nation, 2001) that arguably it seems no longer necessary to emphasise the essential role played by vocabulary¹ in the acquisition of a second language (L2). To put it briefly, we entirely agree with Pérez Basanta (forthcoming) when she contends that “lexis is at the heart of language acquisition”.

This general consensus, however, cannot be extended to the definition of lexical competence. Authors have adopted different standpoints regarding its nature and, as a result, to date no standard approach has been put forward as a generally accepted benchmark of vocabulary knowledge. Undoubtedly, one of the areas where this controversy has very clear implications is lexical assessment since “vocabulary tests are contingent upon the test designer’s definition of lexical knowledge” (Laufer & Goldstein, 2004: 399). From a review of literature, a dichotomy has traditionally been established in the field of vocabulary testing with respect to the nature of lexical competence: the distinction between breadth and depth of knowledge (Anderson & Freebody, 1981). The former attempts to cover the number of words the student knows, i.e. the size of his/her lexicon. By contrast, the latter refers to the degree to which students know words —whether they possess a multidimensional qualitative lexical knowledge including pronunciation, spelling, meaning, register, frequency, and grammatical and collocational patterns (Qian & Schedl, 2004).

These two perspectives have not received, however, equal attention from researchers. Probably due to the fact that it is easier from a practical point of view to test lexical size than depth, measures of vocabulary size are further developed than those of depth (Read, 2000). In this regard, depth tests have often been criticised on the grounds that “the number of items that can be tested is limited and the test does not, therefore, represent the true vocabulary of the test taker” (Laufer & Goldstein, 2004: 401).

However, recent studies have revealed that both dimensions play an important role in language skills (Qian, 2002; Qian & Schedl, 2004). Taking on board these findings, we consider that more research on the assessment of lexical depth is necessary and even urgent. In an attempt to contribute to this field, and taking into account the aforementioned limitation generally attributed to depth tests, we believe that it is not necessary to evaluate the different components of vocabulary depth simultaneously since their different nature makes it possible to assess them individually. In our view, lexical assessment would benefit from the use of

independent measures for each component of lexical knowledge in a number of ways. In terms of test design and scoring, the number of variables involved would be smaller and thus easier to control for the researcher. In addition, it would be possible to test a larger sample of items obtaining, therefore, a more representative and reliable measure. Of course, since the different components of lexical knowledge are intrinsically connected, it would be advisable to later establish statistical correlations between them to obtain a comprehensive estimation of students' lexical competence. Following this assumption and as an initial step in the assessment of L2 learners' lexical depth dimension, we have constructed a test for assessing one of the constituent traits of lexical competence: knowledge of collocations. Of course, ours is not the first test exclusively devoted to measuring collocational competence but, as it will hopefully be made clear in this article, we have tried to develop a reliable and systematic framework which may be used as a starting point for future research on collocational testing. To this end, we will first discuss our theoretical construct and will offer a brief review of previously existing collocational measures, followed by a full account of our study. In this regard, we will explain the procedures taken in terms of test design, administration and statistical analysis of results.

II. COLLOCATIONAL KNOWLEDGE AND PREVIOUS COLLOCATIONAL TESTS

Depending on the perspective adopted, collocations have been approached in different ways by different authors. It is, however, beyond the scope of this article to offer a thorough account of all the existing definitions. Thus, we will only concentrate here on the construct of collocations we considered in our study. To our knowledge, collocations are characterised by a number of formal and functional features. From the formal perspective, they are integrated by two elements: the base and the collocate (Hausmann, 1989). In this regard, we agree with scholars who believe that these two components do not share the same linguistic status since the base is semantically autonomous whereas the collocate is determined and somehow selected by the base (Nesselhauf, 2005).

With respect to functional characteristics, collocations are institutionalised combinations of words which, due to their frequency in the language, have become an integral part of the norm and not only of the system. Thus, given their relative frozenness, collocations are a constituent element of the phraseological inventory of a language. As Pawley and Syder (1983: 209) envisaged some years ago: "What makes an expression a lexical item, what makes it part of the speech community's common dictionary, is (...) that it is a social

institution. This (...) characteristic is sometimes overlooked, but is basic to the distinction between lexicalized and non-lexicalized sequence. (...) Rather than being a ‘nonce form’, a spontaneous creation of the individual speaker, the usage bears the authority of regular and accepted use by members of the speech community”.

However, not all frequent lexical combinations are interesting for L2 learners. Although “open the door” may be a very frequent co-occurring expression in the language, it can be considered a free combination as it can be generated by students simply by applying their grammatical and semantic knowledge. Therefore, we consider that a second functional intrinsic feature of collocations is their arbitrariness. This characteristic is responsible for the fact that “some words are more likely to combine with specific items to form natural-sounding combinations while other types of combinations are simply not found, even though they would be possible and understandable, at least theoretically” (Fontenelle, 1994: 42). For this reason, although the expression “to finish a war” is acceptable as grammatically and semantically correct, a native speaker would usually say “to end a war” following the arbitrary restrictions of the language.

It is widely attested today that collocations play an essential role in SLA. In Lewis’ (2000: 8) words, “the single most important task facing language learners is acquiring a sufficiently large vocabulary. We now recognise that much of our ‘vocabulary’ consists of prefabricated chunks of different kinds. The single most important kind of chunk is collocation. Self-evidently, then, teaching collocation should be a top priority in every language course”. One of the main reasons why collocations are paramount amongst the rest of lexical elements is their aforementioned high frequency in the language. As opposed to idioms, collocations can hardly be paraphrased or substituted by a synonymous expression (Farghal & Obiedat, 1995) so they are essential for the non-native speaker in order to use the language fluently and accurately.

Equally noteworthy is the fact that collocations constitute a problematic aspect for L2 learners. From a purely linguistic perspective, it seems reasonable to assume that the arbitrary nature of collocations is responsible for the attested difficulties of non-native speakers. Moreover, from a more pedagogical approach, different explanations have been put forward to account for this phenomenon. On the one hand, it has been argued that students’ lack of awareness of the existence of collocational patterns results in excessive reliance on L1 to L2 transfer (Farghal & Obiedat, 1995). Thus, students tend to produce deviant collocations following the wrong assumption that there is always a one-to-one correspondence between their mother tongue and the target language in terms of collocations. On the other hand,

authors who are more concerned with psycholinguistic views, contend that the main reason why collocations are a difficult aspect for non-native speakers is to be found in the way they acquire and mentally organise new vocabulary. Contrary to natives, L2 students seem to start by learning individual words and gradually build up bigger chunks, so it becomes particularly hard for them to establish strong associations between pairs of words forming collocations (Schmitt & Underwood, 2004; Wray, 2002). For this reason, they tend to overuse the creative combination of isolated words, rather than store and produce ready-made collocations. Using Sinclair's (1991) terminology, they typically rely on the open-choice principle of language, whereas the idiom principle is often neglected by L2 learners.

Taking into account the previous arguments, one would expect collocations to represent a well-established aspect of vocabulary assessment. Unfortunately, this is not the case. Collocations have been traditionally ignored in the field of language testing, and only in the last few years the need to tap into learners' collocational competence has started to be advocated. To the best of our knowledge, eight projects have been devoted to this area so far²: Biskup (1992), Bahns and Eldaw (1993), Farghal and Obiedat (1995), Bonk (2001), Mochizuki (2002), Barfield (2003), Gyllstad (2005) and Keshavarz and Salimi (2007). In the rest of this section we endeavour to review and compare these studies.

As is the case with the rest of vocabulary tests, collocational measures seem to fall into two categories: the ones which attempt to test productive knowledge and those assessing receptive knowledge. The former was the only aspect contemplated during the decade of the nineties, when Bahns and Eldaw (1993), Biskup (1992) and Farghal and Obiedat (1995) designed the first tests of collocations. Thus, these tests presented the test-taker with a translation task where the target collocations had to be supplied. In addition, Bahns and Eldaw as well as Farghal and Obiedat combined this with a completion format where subjects were also required to fill sentence gaps. The similarities among these three tests can also be found in the limitations they share. In terms of the number of items tested, Biskup does not specify how many collocations were included in her study. In the case of Bahns and Eldaw, their test measured only 15 items, and this number was increased only up to 22 items by Farghal and Obiedat. On the whole, we can observe that the sample of collocations assessed is so small that the conclusions drawn by these studies might be questionable. In addition to this, the unsystematic way in which the specifications have been established in Bahns and Eldaw's test, (in the other two cases not even reported at all), is another matter of concern. Our final criticism has to do with lack of statistical reliability analyses in all the three tests.

Turning our attention to the tests designed in the current decade (Barfield, 2003; Bonk, 2001; Gyllstad, 2005; Keshavarz & Salimi, 2007; Mochizuki, 2002), they all fall into the second category mentioned above since all tap the receptive dimension of collocational knowledge. Special attention is due to Bonk's study, as it is the only one aiming to cover both productive and receptive competence. Nonetheless, it must be noted that he only performed correlation analyses between collocational proficiency and general English proficiency, whereas no internal comparison was established between the receptive and productive dimensions of students' collocational competence. Therefore, this is an aspect still to be tackled in this area of research.

One of the basic notions one perceives after a review of these five recent studies is that the best format to assess collocations receptively seems to be multiple choice items. In fact, this is the format most receptive collocational tests tend to take, and when some other arrangement has been tried out it has been finally rejected in favour of multiple choice items. This was the case of Gyllstad's COLLMATCH test, which was arranged in grids consisting of 3 verbs and 6 nouns, where students were required to indicate the possible combinations existing between them. An example of this test is offered below (Fig. 1). However, due to the attested difficulty in finding nouns that fit with more than one verb, the majority of the combinations in the grid produced deviant collocations. Therefore, "the test primarily measured learners' ability to reject pseudo-collocations (65%), rather than their ability to recognize real collocations (35%)" (Gyllstad, 2005: 22). For this reason, the author had to modify this version and create a multiple choice one.

	charges	patience	weight	hints	anchor	blood
drop						
lose						
shed						

Fig. 1. Example of a COLLMATCH grid, version 1 (Gyllstad, 2005: 16).

It should be noted here that, as opposed to the first tests, recent measures offer more conclusive and reliable findings. This improvement is a consequence of the fact that they use a larger number of items—they range from 50 items in the case of Bonk, 2001 and Keshavarz and Salimi, 2007 to 150 items in Gyllstad's final tests—and they have been subjected to more adequate statistical analyses. There still exists, however, an aspect which continues to be especially problematic in the field of collocational assessment: the selection of items. In the case of Bonk and Keshavarz and Salimi, this selection is notably unsystematic since they seem to take intuition as their only criterion. Moreover, although the rest of the receptive tests show the first attempts of researchers to perform a systematic corpus-based selection of items, they also present some drawbacks. As regards Mochizuki and Gyllstad's studies, the main criterion used to include collocations in their tests was the individual frequency of the words they contained. To make this selection more systematic, Gyllstad also performed a z-score analysis on all the collocations obtained in order to check whether they were all frequent combinations in the British National Corpus. In our opinion, selecting collocations on the basis of the frequency of both words as independent entities reflects the theoretical assumption that the two elements integrating collocations are at the same linguistic level. Contrary to this opinion, however, a number of studies (Corpas Pastor, 1996; Hausmann, 1989; Mel'čuk, 1998) have shown that one of the elements of a collocation is always a semantically independent base which is freely chosen by the speaker whereas the other element is a collocate whose meaning and use is restricted by the base. Following this definition, we believe that the most appropriate way of searching collocations would be to select the base from a frequency word list and then to choose its most frequent collocate according to data obtained from corpora with the help of concordancers, thus following a corpus-driven approach. It is important to remark here that the only test where this design has been put into practice to date is Barfield's, although not without some difficulties. In short, the main problem arising in his research, and a serious one insofar as it may affect the

construct validity of the test, is the inclusion of some items which could be considered as free combinations rather than real collocations (e.g. protect body, govern country, etc.). This issue will be discussed further when dealing with the selection of items in our test.

Finally, three other aspects are worth mentioning to complete this short account of the state of the art in terms of collocational testing. Firstly, most studies have been traditionally devoted to verb-noun combinations due to their high frequency in the language, whereas the rest of the syntactical collocational patterns have been largely neglected. Secondly, studies based on corpus techniques (Barfield, 2003; Gyllstad, 2005; Mochizuki, 2002) have relied on data provided by one corpus, with the limitations that this entails in terms of linguistic representativeness, whereas no attempt has been made to compare data from more than one corpus. Thirdly, as far as we know, no research has been conducted so far to assess the collocational competence of Spanish learners of English.

In the light of this situation, it seems obvious that more research is necessary in order to improve the assessment of L2 students' collocational competence. Hence, the study we present here is an attempt to contribute to this field.

III. THE PRESENT STUDY

In the rest of the paper we intend to spell out the procedures taken in order to construct a reliable and valid test. Consequently, we will report in detail on the decisions taken from a three-fold perspective: 1) Subjects, 2) procedures, and 3) analysis of results. By and large, the research questions we wish to answer in this pilot study are:

- 1) Are Spanish university students able to recognise and produce collocations composed of highly frequent nouns and their most frequent adjectival collocates?
- 2) Is there a significant difference among students' performance in a receptive and a productive collocational measure?

III.1. Subjects

Our test was administered to a total of 63 students at the University of Granada. They were all on their second year of degree of English Linguistics so they had already attended a number of compulsory subjects devoted to improving their general English proficiency. Candidates' age ranged from 19 to 23 and there were 51 women and 12 men. Their native language was Spanish in all cases except for one student whose first language was Ukrainian. As one of the aims in our study was to obtain generalisable data about the collocational competence of

Spanish speakers, the scores yielded by this learner were not used in the analysis of results. This left, therefore, a sample of 62 subjects for this first pilot administration of our test.

III.2. Procedures

To ensure the two fundamental requirements which any test should meet, i.e. validity and reliability (Henning, 1987), a careful selection of contents and construction of items was essential. It is a truism to say that a test has validity when it measures what it is intended to measure and, as far as this test is concerned, we want it to measure collocations found in real and authentic language. In fact, it is natural to think that students “need to learn language as it is used by native speakers for real purposes, rather than language “invented” by linguists and textbook writers” (Baddock, 1996: 20). Hence, the most appropriate tools test designers currently have at their disposal to establish test validity are computerised corpora, insofar as they provide them with authentic and representative samples of the language. For more than a decade now, electronic corpora have been considered a resource for providing authentic language exposure (Pérez Basanta & Rodríguez Martín, 2006). Thus, authenticity is a notion very often associated with corpus-based teaching in general and concordances in particular, inasmuch as “you can present your students with *authentic* evidence from which they can work out answers for themselves” (Tribble & Jones, 1990: 10). Consequently, we advocate a corpus-driven approach for the selection of test contents which eventually will be the material for the design and elaboration of items as a guarantee of test validity.

In terms of reliability of the test —we all know that a test is reliable if it consistently gives the same results under different conditions— perhaps the most important sources of consistency are: item construction, scoring and administration. In the next sections, we will explain how we have tried to ensure the test met the criteria of validity and reliability.

III.2.1. Selection of items

First, it should be noted here that all the collocations selected in our study were adjective-noun structures since, in our opinion, it is a pattern which deserves more attention as it has traditionally been neglected in previous studies.

As aforementioned, following the assumption that collocations are integrated by two unequal elements, the selection procedure carried out consisted of two main processes: the search for the bases among the most frequent English nouns, followed by the selection of their

most frequent adjectival collocates drawn from corpus data. Thus, the following steps were taken to ensure content validity, that is to say, a good selection of test contents.

To accomplish the first stage, we examined the 1,000 most frequent words of a frequency list developed under the umbrella of the ADELEX R&D project undertaken at the University of Granada. The ADELEX word count contains the 7,124 most frequent words of English and it is the result of a number of comparative analyses of data provided by the most reliable and representative English corpora currently available: the Bank of English, the British National Corpus and the Longman Corpus Network (López-Mezquita, 2005). Although the final number of selected items was 80, such a wide margin of words was necessary (up to 1,000) since, on the one hand, the only valid word class for our purpose were “nouns” and, on the other, only those nouns which formed frequent collocations with an adjective were interesting for us.

To confirm whether a base was acceptable or not, we proceeded to the second phase of content selection: the search for collocates. To this end, a list of the most frequent words co-occurring with each noun was first drawn from the British National Corpus (BNC); this was undertaken by running the “show collocates” function of the program *Concord* from *WordSmith Tools* (v. 3.0), and sorting the resulting list by left. Since corpus-based software can only perform statistical calculations, the results provided need to be manually scrutinised in order to identify and reject frequent combinations which, according to our theoretical construct, might not be interesting collocations for L2 learners. From our experience, and we believe it was also the case in Barfield’s study when running a similar search, this is the most complex aspect when performing a corpus-driven selection of collocations as very clear criteria need to be established to discriminate between those collocations which may be necessary for our students (and, therefore, should also be included in a test of collocations) and those which are not useful despite their high frequency. In our case, we considered interesting those adjective-noun collocations which met the following criteria: 1) they were used in a wide range of texts types and contexts (very specialised or subject-dependent collocations such as “single market” were rejected), 2) they were semantically transparent (as opposed to expressions such as “bottom line”, classified as idioms given their non-compositionality), and 3) they were arbitrarily restricted in their commutability and/or combinability being, thus, necessary for learners to store them as single units (as opposed to free combinations of words which can be freely built by speakers simply by applying the rules of semantics and grammar, such as “young people”). Evidently, we advocate an eclectic

selection procedure where collocations are driven from corpus data, but where the arbitrary restrictions imposed by their phraseological nature are also considered a discriminating factor.

Once the most frequent collocates were chosen, the next step was to check the frequency of the whole combination in the BNC using a span of 4 words (Jones & Sinclair, 1974) to the left of the noun. Ten instances of the collocation were considered the minimum level of acceptance³, although most of them proved to be much more frequent.

In order to improve the validity of our measure, the final step in the selection process was contrasting the data offered by the BNC with information from the Bank of English (BOE). From our viewpoint, this is a particularly noteworthy aspect since, as compared to the rest of the collocational measures we are aware of, this is the first time that the selection of a bank of collocations bears upon a comparative and contrastive analysis of data from two different corpora. It seems clear to us that this double-check procedure is a further guarantee for ensuring content validity. Using the online Collins Cobuild Corpus Collocations Sampler, T-score and Mutual Information analyses were performed for each base word. Although on some occasions the data provided by the BNC coincided with that from the two measures based on the BOE, it was very common to find collocations rendered as very frequent by WordSmith and T-score programs but which did not seem to be significant when results from Mutual Information analyses were examined. Nevertheless, these combinations were accepted for inclusion in our test as it was considered that the reason why the Mutual Information scores were not significant was the high frequency of both words as independent units. One final case to highlight here are the few combinations which were significantly frequent collocations according to data from the BOE but which did not appear at all in the BNC. In most of the cases these combinations were included in the test as it was considered that this incongruity was a consequence of the inherent differences existing between both corpora. In this sense, since the BOE is a monitor corpus which is continually updated with new texts, it contains collocations which are very frequent nowadays but which were not common before the nineties, when all the texts included in the BNC were collected.

The final result of this exhaustive selection procedure was a bank of 80 frequent collocations, a sample which was considered large enough for collocational assessment.

III.2.2. Test construction, administration and scoring

Undoubtedly, reliability bears upon test construction, administration and scoring. It is widely-held that to increase reliability of tests, considerable efforts must be made by the testers to specify explicitly the tasks to be performed, define the criteria against which performance is

to be assessed and establish a score which is as objective as possible. Therefore, an objective 80-item test was designed, divided into two 40-item sub-tests, one for receptive knowledge and the other for productive knowledge. The multiple choice format was selected for the Receptive Collocational Test (henceforth RCT) given the objectivity of scoring it allows. Students were presented with the definitions of the concepts expressed by the target collocations as provided by the *Collins COBUILD Advanced Learner's English Dictionary on CD-Rom* (2003). The dictionary was a valuable asset in this case as it guaranteed the necessary correction and accuracy of the item prompts. An example of an item from the RCT is provided below.

A place where it is unlikely that any harm, damage or unpleasant things will happen to the people or things that are there can be called a...

- a. security place b. sure place c. safe place d. none of these

As suggested by Gyllstad (2005), a basic criterion for the construction of item options was the need to create tempting pseudo-collocations which would seem plausible as an alternative to real collocations. To this end, we relied on intuition and teaching experience about interferences between L1 and L2 collocations. Nevertheless, a further corpus-based analysis was performed on distractors as they were all checked in the BNC to certify that they were not acceptable combinations in English. Finally, as it can be observed in the example above, the fourth option provided in every item was “none of these”. This alternative, which was the correct answer in 10% of the items, was introduced to minimise the effect of guessing (López-Mezquita, 2005). Since the test-taker cannot concentrate on a number of alternatives one of which is necessarily correct, this option improves test discrimination and reliability.

For the assessment of candidates' productive collocational knowledge, another 40-item test was constructed. Here we opted for another discrete-point item type: gap-filling. In this case, this item-response format was closed-ended and students were asked to complete a definition of the concept expressed by the intended collocations, in other words, they were requested to supply the adjective, i.e. the collocate, which is constrained in its commutability and/or combinability by the noun given. When these items prompted more than one correct answer, they were all accepted. This was, for example, the case in the following item from the PCT, where both “sound argument” and “strong argument” were accepted:

When you have good reasons to support or oppose an idea or suggestion you have a.....
ARGUMENT

Once the test was finished, the two parts were administered together in pencil-and-paper format. Two further points might prove of interest. In terms of administration and timing, students were allowed 60 minutes to complete the test although most of the subjects were able to finish it before that time, indicating that the measures were correctly designed from a practical point of view. Another thing to consider was place. The test was administered in the same classroom and students were separated as much as possible to reduce any cheating to a minimum. Finally, the third source of reliability was test scoring. In fact, it was a straightforward task given the fact that, as it was previously mentioned, items were designed in objective formats. There was no problem of inter-rater reliability as I myself corrected the whole test. Correct answers scored one point and incorrect or badly spelt answers (in the PCT) scored zero.

Once the test was marked, we gathered the resulting data and produced a database using *The Statistical Package for Social Sciences* (SPSS 14) which helped us to summarise descriptive statistics (central measures and dispersion) and carry out some inferential statistics based on t-tests.

III.3. Results and discussion

Our first research question concerned Spanish university students' general knowledge of frequent collocations. Figure 2 shows that the total mean of correct answers in the whole test (i.e. including both sub-tests) was 38.3%, a considerably low score. Furthermore, the relatively low standard deviation (S.D.) (10.36) shows that the group is fairly homogeneous in their level of collocational knowledge. From these data we tentatively conclude that the overall collocational competence of our students falls short of our expectations in this respect.

	Total (RCT + PCT)	RCT	PCT
N	62	62	62
Mean	38,3065	46,0887	30,5242
S.D.	10,36994	12,39398	10,95012
Variance	107,536	153,611	119,905
Range	53,75	65	52,5
Min.	11,25	15,00	7,50
Max.	65,00	80,00	60,00

Fig. 2. Descriptive statistics for collocational tests in percentages.

The second research question addressed in this pilot study asked whether there was a significant difference among learners' scores on the receptive and productive collocational tests. From a comparison of the data obtained from both sub-tests (Fig. 2), we observe a clear difference between the mean scores in the RCT (46.08%) and the PCT (30.52%) —although it is also noticeable that neither of them reaches 50%, the minimum figure to consider that the test was “passed”. A t-test of these two means confirmed that the difference between them was highly significant ($p=.000$) and, thus, not due to chance. In the light of these results we can conclude that collocations have proven to be more difficult at the productive than at the receptive level, a finding which empirically confirms the generally held hypothesis that this type of combination is particularly problematic for students in their linguistic production. Equally interesting is the information concerning S.D. in both sub-tests. Oddly enough, subjects' scores were more uniform in the PCT (10.95) than in the RCT (12.39). One possible explanation for this could be that the RCT discriminates between high and low level candidates while the PCT produced such low scores for most students that no variance is observable: all candidates show the same lack of knowledge. This is also supported by the range, which amounts to 65% in the RCP as compared to 52.5% in the PCT.

In order to obtain the reliability coefficient, we ran a Cronbach's alpha (α) analysis for total scores and for sub-tests individually. The internal reliability values found in the RCT ($\alpha .699$) as well as in the PCT ($\alpha .736$) were relatively acceptable, though not completely satisfactory as they did not reach .8 —a conventional yardstick against which reliability is measured (Abad et al., 2004). We consider, however, that these moderately low scores were due to the small number of items (40 each) and the little variance existing among subjects' performance. But taking the test as a whole, its overall reliability was highly acceptable (α

.818). This satisfactory result may be attributable to the careful and systematic corpus-based design and, perhaps, to the construction of the test items described above.

Finally, a thorough item analysis was conducted in order to obtain the index of item difficulty⁴ and the index of item discrimination⁵, and ultimately deciding which items should be accepted or rejected for an optimisation of our pilot test. As shown in Figure 4, after an analysis of item difficulty, 5 items (all of them belonging to the PCT) obtained p-values of .0 since they prompted incorrect answers from all the participants. As expected, the discrimination index (obtained by performing a point-biserial correlation⁶) showed that these highly difficult items were non-discriminating among candidates, and so they would need to be replaced in future studies by more relevant items. The rest of the numerical values yielded by the item difficulty analysis were classified following Ebel's (1965, cited in Cervantes, 1989) criteria (Fig. 3): 15 items (18.75%) were classified as very difficult, 24 items (30%) as difficult, 23 items (28.75%) offered a desirable level of difficulty, 6 items (6.5%) were easy and finally 7 items (8.75%) fell into the category of very easy items. From our point of view, the most outstanding consequence to be drawn from this in-depth item analysis is the corroboration of the previous finding concerning the poor collocational competence of Spanish university L2 students. In addition, taking into account the distribution of the values between the RCT and the PCT tests, the results obtained are further good evidence of the fact that producing collocations is a far more complex task for L2 learners than recognising them.

	Overall	RCT	PCT
Non-discriminating items (p-value = .0)	5 (6.25%)	0 (0%)	5 (6.25%)
Very difficult items (p-values from .01 to .14)	15 (18.75%)	4 (5%)	11 (13.75%)
Difficult items (p-values from .15 to .39)	24 (30%)	14 (17.5%)	10 (12.5%)
Desirable items (p-values from .40 to .70)	23 (28.75%)	14 (17.5%)	9 (11.25%)
Easy items (p-values from .71 to .85)	6 (7.5%)	3 (3.75%)	3 (3.75%)
Very easy items (p-values from .86 to 1)	7 (8.75%)	5 (6.25%)	2 (2.5%)

Fig. 3. Analysis of difficulty of individual items according to Ebel's criteria.

Some examples of items incorrectly answered by all test-takers are “half term” and “wrapping paper”. Despite their frequency, we construe that these are collocations very much related to the culture and everyday life of English speaking countries and this may be the reason why they are especially problematic for L2 students. On the other hand, there were 7 conspicuously easy items, 5 in the RCT and 2 in the PCT. Especially interesting is the collocation “Prime Minister” included in the latter, which was the only one prompting over 90% correct answers. The main reason why this may be such an easy collocation for our students is that it is constantly present in media. We would like to note that we are fully aware of the fact that a deeper and more detailed analysis of the linguistic and psycholinguistic factors affecting the intrinsic difficulty of collocations would be necessary here. Due to space constraints, however, this complex issue will be addressed in future studies.

IV. CONCLUSION

This paper has suggested that, in order to assess students' depth of lexical competence, it would be advisable to design individual measures for each component of the multifaceted nature of vocabulary. Our study is an attempt to delve into the area of collocational assessment, where, given the limitations of the few contributions made in this field to date, further research is warranted.

A careful revision of literature on collocational testing reveals that until very few years ago, one of the main weak points of collocational tests was their lack of systematicity. A generalised reliance on human intuition was the basis for their design and construction stages, an approach which led to highly invalid results. Aware of this, and in an attempt to surpass these traditional limitations in tests' content validity, measures designed in the 21st century have propounded a corpus-driven approach to collocational testing. In our view, however, this has only been partially attained so far since, although frequency lists have provided a new computational and scientific basis for the selection of frequent words co-occurring in collocations, in most of the cases target collocations are not “driven from” but “based on”⁷—i.e. checked with— corpus data. In the present paper an attempt has been made to put forward a framework to measure collocational competence by tapping into students' knowledge of frequent collocations as integrated by a frequent base and its most frequent collocate, irrespective of the frequency of the latter as an isolated word. To us, if the natural

features of collocational patterns are to be taken into account, more efforts should be devoted to an in-depth examination of what information corpora can provide.

We hope to have made clear, however, that we do not aim at a radical statistically-based extraction of collocations. Given the phraseological nature of collocations, we advocate an eclectic approach. On the one hand, we should benefit from the advantages of corpus-based methodologies and empirical data, but, on the other hand, human experience should also be given its rightful place as the necessary element to judge the inherent arbitrariness which characterises and ultimately determines the nature of collocations, as opposed to free combinations and idioms.

On the whole, the results yielded by our pilot test lead us to conclude that the overall collocational competence of students of English Linguistics at the University of Granada is insufficient and this indicates that students may fall short in the social and academic demands made on their command of L2. Moreover, this research study has also confirmed the widely held notion that collocations are more problematic from a productive perspective (Hussein, 1990). The major implication of these results, moreover, would seem to be the urgent need to carry out an efficient pedagogical intervention to overcome students' collocational deficiencies.

To conclude, a further note should be added concerning our limitations and ways forward. Firstly, the study presented was only administered to second year university students. In this respect we consider that, in order to measure the collocational knowledge of our students, a longitudinal assessment comparing the performance of candidates from different university levels would be advisable. Secondly, we are aware of the fact that the results yielded may not be generalisable to a large population given the relatively limited number and scope of subjects participating in this pilot study. Nevertheless, we believe that a replicable framework has been put forward for similar tests to be designed in different contexts. Finally, the statistical analysis of items carried out has shown that some elements do not offer discriminating information from test-takers. Clearly, these items need to be replaced in future studies in order to optimise results.

It goes without saying that there is still a good way ahead to offer conclusive evidence of the design of a definite test to assess collocational knowledge, but hopefully this paper can help to provide the direction and focus area for further research.

Acknowledgements

I would like to thank Dr. Carmen Pérez Basanta and Dr. Elizabeth Ellis for their helpful insights and comments on a previous draft of this paper. This research was conducted under the auspices of the research grant FPU (AP2005-1353) and within the R&D project entitled “ADELEX: Assessing and Developing Lexical Competence through New Technologies” (Ref: HUM2007-61766), both financed by the Spanish Ministry of Education and Science.

NOTES

1. Throughout this paper we use the term “vocabulary” and “lexis” interchangeably.
2. In the concise review we offer here we concentrate our attention on those tests which have been designed to evaluate collocational competence exclusively. Therefore, we won't address other measures which include collocations as one of the components of the wider lexical construct being assessed (see, for example, Qian, 2002 or Boers et al., 2006).
3. The threshold used results from dividing the square root of the corpus size by 1,000,000 (Mason, 2006).
4. This measure (also called *p-value*) calculates the proportion of subjects answering correctly to each item. It can range between 0.0 and 1.0, with a higher value indicating that a greater proportion of examinees responded to the item correctly, and it was thus an easier item.
5. This is a measure of how well the item discriminates between skilled and unskilled examinees. It usually correlates with the index of test difficulty because those items which are very easy or very difficult do not discriminate between candidates who are knowledgeable in the content area and those who are not.
6. The point-biserial correlation looks at the relationship between an examinee's performance on a given item (correct or incorrect) and the examinee's score on the overall test (Henning, 1987).
7. We follow Tognini Bonelli's (1996) terminology to establish a distinction between corpus-driven and corpus-based patterns.

REFERENCES

- Abad, F. J., Atencia, J., García C., Hontangas, P., Olea, J., Ponsoda, V., Revuelta, J., Suero, M. & Ximénez, C. (2004). *Ayuda a la creación de exámenes*. Available at <<http://www.uam.es/docencia/ace>> [retrieved 26/05/2007]
- Anderson, R. & Freebody, P. (1981). Vocabulary knowledge. In J.T. Guthrie (Ed.), *Comprehension and teaching: Research reviews*. Newark, DE: International Reading Association, pp. 77-117.
- Baddock, B. (1996). *Using films in the English classroom*. Hertfordshire: Phoenix Ltd.

- Bahns, J. & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21:1, 101-114.
- Barfield, A. (2003). *Collocation recognition and production: Research insights*. Tokyo: Chuo University.
- Biskup, D. (1992). L1 influence on learners' renderings of English collocations. A Polish/German empirical study. In P. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics*. London: Macmillan, pp. 85-93.
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H. & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, 10:3, 245-261.
- Bonk, W.J. (2001). Testing ESL learners' knowledge of collocations. In T. Hudson & J.D. Brown (Eds.), *A focus on language test development: expanding the language proficiency construct across a variety of tests. (Technical Report #21)*. Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center, pp. 113-142.
- Bogaards, P. & Laufer, B. (Eds.) (2004). *Vocabulary in a second language*. Amsterdam: John Benjamins.
- Carter, R. (1987). *Vocabulary. Applied linguistic perspectives*. London: Routledge.
- Cervantes, E. (1989). Designing a reading and listening test for a specific purpose. *English Teaching Forum*, 27:1, 10-14.
- Collins COBUILD Advanced Learner's English Dictionary on CD-Rom*. (2003). London: HarperCollins Publishers.
- Corpas Pastor, G. (1996). *Manual de fraseología española*. Madrid: Gredos.
- Farghal, M. & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *International Review of Applied Linguistics*, 33:4, 315-331.
- Fontenelle, T. (1994). What on earth are collocations? *English Today*, 10:4, 42-48.
- Gyllstad, H. (2005). Words that go together well: Developing test formats for measuring learner knowledge of English collocations. In F. Heinat, and E. Klingvall, (Eds.). *The department of English in Lund: Working papers in linguistics*. Vol. 5, pp. 1-31. Available at <<http://ask.lub.lu.se/archive/00022547/01/gyllstad-wp-05.pdf>> [retrieved 20/05/2007].
- Hausmann, F. (1989). Le dictionnaire de collocations. In F. Hausmann, H. Wiegand & L. Zgusta (Eds.), *Wörterbücher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie*. Berlin: de Gruyter, pp. 1010-1019.
- Henning, G. (1987). *A guide to language testing: Development, evaluation, research*. Cambridge, Mass.: Newbury House.

- Hussein, R. (1990). Collocations, the missing link in vocabulary acquisition amongst EFL learners. *Papers and Studies in Contrastive Linguistics*, 26, 123-136.
- Jones, S. & Sinclair, J. (1974). English lexical collocations. *Cahiers de Lexicologie*, 24, 15-61.
- Keshavarz, M. H. & Salimi, H. (2007). Collocational competence and cloze test performance: a study of Iranian EFL learners. *International Journal of Applied Linguistics*, 17:1, 81-92.
- Laufer, B. & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54:3, 399-436.
- Lewis, M. (1993). *The lexical approach*. Hove: Language Teaching Publications.
- Lewis, M. (2000). Introduction. In M. Lewis (Ed.), *Teaching collocation. Further developments in the lexical approach*. Hove: Language Teaching Publications, pp. 8-9.
- López-Mezquita Molina, M. T. (2005). *La evaluación de la competencia léxica: Tests de vocabulario. Su fiabilidad y validez*. Madrid: Ministerio de Educación y Ciencia.
- Mason, O. J. (2006). *The automatic extraction of linguistic information from text corpora*. Unpublished Doctoral Dissertation, University of Birmingham, United Kingdom.
- McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.
- Mel'čuk, I. (1998). Collocations and lexical functions. In A. Cowie (Ed.), *Phraseology. Theory, analysis, and applications*. Oxford: Oxford University Press, pp. 23-53.
- Mochizuki, M. (2002). Exploration of two aspects of vocabulary knowledge: Paradigmatic and collocational. *Annual Review of English Language Education in Japan*, 13, 121-129.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.
- Pawley, A. & Syder, F. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. Richards & R. Schmidt (Eds.), *Language and communication*. London: Longman, pp. 191-226.
- Pérez Basanta, C. Forthcoming. A second-generation CALL vocabulary-learning programme (ADELEX): In search of a psychopedagogic model. *Estudios de Lingüística Inglesa Aplicada*, 7.
- Pérez Basanta, C. & Rodríguez Martín, M.E. (2006). The application of data-driven language to a small-scale corpus: using film transcripts for teaching conversational skills. In E.

- Hidalgo, L. Querada & J. Santana (Eds.), *Corpora in the foreign language classroom*. Amsterdam: Rodopi, pp. 141-160.
- Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52:3, 513-536.
- Qian, D. & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, 21:1, 28-52.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Schmitt, N. & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins, pp. 173-190.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Tribble, C. & Jones, G. (1990). *Concordances in the classroom: A resource book for teachers*. London: Longman.
- Tognini Bonelli, E. (1996). *Corpus theory and practice*. Birmingham: Tuscan Word Centre.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Zimmerman, Ch. B. (1997). Historical trends in second language vocabulary instruction. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition. A rationale for pedagogy*. Cambridge: Cambridge University Press, pp. 5-19.