

The Use of Collocations by Advanced Learners of English and Some Implications for Teaching

NADJA NESSELHAUF

Universität Basel

Although it is widely acknowledged that collocations are both indispensable and problematic for language learners and that they therefore should play an important part in second language teaching, especially at an advanced level, learners' difficulties with collocations have not been investigated in much detail so far. This paper reports on an exploratory study that analyses the use of verb–noun collocations such as *take a break* or *shake one's head* by advanced German-speaking learners of English in free written production. First, an attempt is made to define 'collocations' as precisely as possible, and the methodology that has been developed for analysing learner collocations in free production is described. Then, the types of mistakes that the learners make when producing collocations are identified and the influence of the degree of restriction of a combination and of the learners' L1 on the production of collocations is investigated. While the degree of restriction emerges to have some, but comparatively little, impact on the difficulty of combinations for the learners, the learners' L1 turns out to have a degree of influence that goes far beyond what earlier (small-scale) studies have predicted. Finally, the implications of these results for teaching are discussed, most importantly the role of L1–L2 differences.

1. INTRODUCTION

That collocations, that is, word combinations such as to *make a decision* or a *bitter disappointment*, are an important part of native speaker competence, and that they therefore should be included in foreign and second language teaching is widely acknowledged today (e.g. Kennedy 1990; Cowie 1992; Bahns 1997; Granger 1998). Collocations are of particular importance for learners striving for a high degree of competence in the second language, but they are also of some importance for learners with less ambitious aspirations, as they not only enhance accuracy but also fluency (e.g. Wray 2002). However, although some suggestions on the teaching of collocations have been made in recent years (e.g. Lewis 2000), it is largely unclear how and especially which of the great number of collocations in a language should be taught. To answer these questions satisfactorily, it is essential to identify the problems that learners have in dealing with collocations. Owing to the nature of collocations (i.e. the fact that they are fairly transparent), comprehension is

normally unproblematic for learners, so that identifying the problems of learners must mean analysing their production of collocations.

Unfortunately, studies that analyse non-native speaker collocation production are not only rare, but also mostly unsatisfactory. Most of them rely on small-scale elicitation tests, often consisting of translation tasks, making it questionable whether the results are generalizable and whether they reflect the learners' actual production problems (e.g. Gabrys-Biskup 1990, 1992; Bahns and Eldaw 1993). Additionally, in some of these studies, the concept of collocation remains hazy, so that besides combinations such as *heavy drinker* combinations and lexical items such as *striped shirt*, *alarm clock*, or *safety belt* are, without further discussion, included among the items tested (e.g. Hussein 1990; Farghal and Obiedat 1995). So far, only very few studies have analysed learner collocations on the basis of a reasonable amount of natural production data (Chi *et al.* 1994; Howarth 1996; Granger 1998; Lorenz 1999).

The present paper attempts to shed some light on the problems of advanced learners of English in the production of collocations. It reports on a study of verb-object-noun collocations (such as *take a picture* or *draw up a list*) in 32 essays written by German-speaking advanced learners of English.¹ The study is exploratory to a large degree, and, due to the lack of previous studies in the field, a definition of collocations and a methodology that allows their analysis in a corpus of learner language had to be developed. In addition, although the database is larger than in most of the previous studies, it is still comparatively small for firm conclusions. In the first part of the paper, I will discuss the definition of collocations and the methods I have developed. In the second part, the types of mistakes learners make when producing collocations will be identified, and the influence of the degree of restriction of a combination and of the learners' first language on the production of collocations will be investigated. In the last part of the paper, some tentative suggestions will be made as to the implications for the teaching of collocations if the results of this study are borne out by further analysis.

2. DEFINING COLLOCATIONS

In this paper, the term 'collocation' is used in a phraseological rather than in a frequency-based sense, which means that it is used to denote a type of word combination (like, for example, in Cowie 1994) rather than the co-occurrence of words in a certain span (like, for example, in Sinclair 1991). In order to determine the difficulties learners have with collocations, a fairly clear delimitation of collocations from other types of word combinations is necessary and will be attempted here. It has to be pointed out from the beginning, however, that word combinations are not in fact clearly delimitable and any attempt to do so involves both theoretical and practical problems. Different researchers have used different criteria—sometimes simultaneously—to delimit collocations from other types of word combinations, but only the most widely accepted defining criterion for collocations is

used here, namely arbitrary restriction on substitutability. This means that a distinction is made between combinations in which a possible restriction on the substitutability of elements is due to their semantic properties (these combinations will be called ‘free combinations’, following Cowie 1994) and combinations in which this restriction is to some degree arbitrary (i.e. ‘collocations’). For example, in the combination *read a newspaper*, the reason that substitutions resulting in combinations such as **drink a newspaper* or **read water* are not possible or at least highly unusual is that *drink* requires a noun with the semantic property of ‘liquid’ and *read* requires a noun with the semantic property of ‘containing written language’. In the combination *reach a decision*, on the other hand, *decision* can be substituted by a number of nouns denoting ‘a particular aim’ (OALD²) such as *conclusion*, *verdict*, *compromise*, or *goal* but not, for example, by *aim*; this restriction does not seem to be a result of the semantic properties of the two elements concerned, but a somewhat arbitrary convention of the language. This distinction between semantically motivated restriction and arbitrary restriction is, of course, not a rigid one and is often difficult to draw.

The distinction between these two types of restriction is usually even impossible to make with words that only combine with very few other words. Consider the combination *dial a number*. It is hardly possible to determine whether the meaning of *dial* is so specific that it can only be used with the noun *number* or whether there is arbitrary restriction at work, which prevents the combination of *dial* with, say, *radio station* (although the object that is used to carry out the corresponding action can be called a *dial*). To tackle this problem, I have developed a notion which is central to my definition of collocations and thus to my classification of combinations; this notion I would like to call ‘restricted sense’. This notion and the classification based on it have been developed and will be presented for verb–object–noun combinations only (although it should be possible to use them for the classification of other grammatical types of combinations without major modification). A sense of a verb (or noun) is considered ‘restricted’ if at least one of the following criteria applies:

Criterion 1

The sense of the verb (noun) is so specific that it only allows its combination with a small set of nouns (verbs).

Criterion 2

The verb (noun) cannot be used in this sense with all nouns (verbs) that are syntactically and semantically possible.

The primary sense of *want* (‘to have a desire or wish for sth’; OALD), for example, would be considered unrestricted according to this definition, since *want* can be combined with a great number of nouns (*want toys, a child, a drink, a car, truth, etc.*) and there are no arbitrary constraints on its combinability. Similarly, as already indicated above, the verb *read* in *read a newspaper* would be considered unrestricted, since it can be combined (in the

given sense of 'look at and comprehend the meaning of (written or printed matter) by interpreting the characters or symbols of which it is composed'; *CCED*³) with all nouns denoting written or printed matter (*read a book, a letter, a sign, a manual, a leaflet*, etc.). *Dial* or *perform* (as in *perform a task*), on the other hand, would be considered as having restricted senses: *dial*, because it can only combine with one (or at most very few) nouns, and *perform*, because, although it is combinable with a greater number of nouns, some nouns that seem to be possible from a semantic point of view are not possible (e.g. **perform a survey*; cf. Cowie 1994: 3169). An example of a noun with a restricted sense is *pill* in the expressions *sugar* or *sweeten the pill*. The 'sense' of *pill* in these combinations ('unpleasant news or unpleasant measure'; *CCED*) cannot be used freely: **I was confronted with a pill yesterday*, for example, is not possible.

On the basis of this notion of restricted sense, three major classes of word combinations can be distinguished:

Free combinations (e.g. *want a car*):

The senses in which the verb and the noun are used are both unrestricted, so they can be freely combined according to these senses.

Collocations (e.g. *take a picture*):

The sense in which the noun is used is unrestricted, but the sense of the verb is restricted, so that the verb in the sense in which it is used can only be combined with certain nouns (*take a picture/photograph*; but e.g. **take a film/movie*).

Idioms (e.g. *sweeten the pill*):

Both the verb and the noun are used in a restricted sense, so substitution is either not possible at all or only possible to an extremely limited degree.

A combination is thus classified as a collocation if either criterion 1 or criterion 2 or both apply to the verb of the combination. What is new about this definition is not only that it tries to be as precise as possible in order to be applicable to the classification of naturally occurring data, but also that it makes the verb the basis for the distinction of free combinations and collocations, and the noun the basis for the distinction of collocations and idioms: if the sense in which the verb is used is restricted, the combination is a collocation (rather than a free combination), if the sense in which the noun is used is restricted, the combination is an idiom (rather than a collocation). This aspect of the definition is not only supported by its feasibility in the data analysis but also by a theoretical assumption that has been put forward by some researchers—the assumption that the elements in a collocation do not have the same status (e.g. Hausmann 1989, Mel'čuk 1998). It is assumed that one element in a collocation is selected purely on the basis of its meaning, while the selection of the other depends on this first element; it is also assumed that this relationship is consistent in a given syntactic combination, and that in all verb–noun collocations the independent element is the noun and the dependent one the verb. Although this assumption looks at the

relation between the verb and the noun from a different perspective than the definition above, its effects are the same, namely that nouns behave in the same way in free combinations and collocations, and verbs do not. Despite the fact that the distinction of collocations from other types of verb–noun combinations is based exclusively on the verb and the noun in the present definition, however, the term ‘collocation’ will be used to refer to whole combinations in this paper, i.e. including noun pre- and post-modification patterns of the noun (for example, *take pride in* and not only *take pride* will be considered a collocation).

Despite the attempt to be as precise as possible, the classification presented above has a number of limitations, and, as has already been indicated, cannot be without them, as word combinations differ along a scale, which makes their exact delimitation impossible. Therefore, the problem of distinguishing arbitrary and semantically motivated restriction could only be reduced but not solved with the notion of ‘restricted sense’, and the line between collocations and idioms, although less difficult to draw, similarly is not rigid. It is also difficult to determine whether a combination ‘exists’ in a language or not. In this study, combinations are considered to ‘exist’ if they are fairly well established in a language. If a certain combination, say *run a chance*, occurs once or twice in a large corpus, but native speakers consider it unacceptable, it is not taken into account when determining whether and to what degree the verb is restricted. A further problematic point in the definition is the expression ‘small set’ used in criterion 2, as an arbitrary limit has to be set; for this study, up to five elements were considered a small set.

In the remaining part of the paper, free combinations will be abbreviated F, collocations RC (for ‘restricted collocation’, to avoid confusion with the abbreviation for ‘correct’), and idioms I.

3. METHODOLOGY

The investigation into the difficulties of advanced learners with verb–noun collocations in the sense indicated above and into some factors that possibly contribute to these difficulties is based on data from the German subcorpus of ICLE (The International Corpus of Learner English).⁴ The 32 essays investigated were written by German-speaking university students of English, mainly in their 3rd or 4th year, i.e. by learners most of whom are probably trying to achieve a high level of competence in English. The essays are argumentative and non-technical, including essay titles such as ‘Is there any point in being ecology-conscious?’ or ‘Is peace a good thing at any price?’ and have an average length of about 500 words. The selection of the essays from ICLE was largely random, although a few non-linguistic criteria were applied.⁵ The first step in the analysis of the data was to manually extract all verb–object–noun combinations from the essays.⁶ In a second step, the combinations were classified as to their degree of restriction (i.e. F, RC, or I), and in a third step they were evaluated as to their acceptability in English. While the

first step does not require any explanation, step two and step three are now going to be described in some detail.

The second step of the analysis, the application of the definition developed above to actual data, proved to be a great challenge. What was particularly difficult was to find out whether words, especially verbs, were used in a restricted sense in a given combination. This problem is linked to a more general one, namely that native-speaker norms seem to be particularly variable in the area of word combinations: speakers differ in what they find acceptable, individual speakers are often uncertain about whether they find a combination acceptable or not, and, additionally, there can be differences between the acceptability judgements of speakers and their own use. Since it was not possible to carry out corpus analyses for each verb (and noun) found in the learner data in order to determine with what nouns (and verbs) it can occur (the main reason being that manual sense disambiguation would have been necessary), a more pragmatic solution had to be found. The only feasible solution that emerged was the use of dictionaries, combined with some corpus analysis and some native speaker judgements. After comparing a number of dictionaries, I decided to use the *Oxford Advanced Learner's Dictionary (OALD 2000)* and the *Collins COBUILD English Dictionary (CCED 1995)* as the main sources to determine whether a word is used in a restricted sense or not. These two dictionaries were chosen because they were found to give the most detailed information on what words the headwords can combine with.⁷

The procedure was as follows (since nouns were largely unproblematic to classify, this is exemplified with verbs): if the two dictionaries gave no indication of a possible restriction of the verb or sense of the verb in question, i.e. if no specific nouns were given in the definition and there were no other indications of restriction (such as individual examples separated by slashes in the *OALD*, e.g. *take a vote/poll/survey*), the verb was considered as having an unrestricted sense and the combination was classified as free (F).

For example, *need*:

OALD: to require sth/sb because they are essential or very important, not just because you would like to have them

CCED: If you **need** something, or **need** to do something, you cannot successfully achieve what you want or live properly without it.

If there were clear indications that the use of the verb (in a particular sense) is restricted to a few nouns, it was considered as having a restricted sense in the context, and the combination was classified as a collocation (RC).

For example, *fail*:

OALD: TEST/EXAM to not pass a test or an exam

CCED: If someone **fails** a test or examination, they perform badly in it and do not reach the standard that it required.

If the signals were unclear, that is if, for example, the definition included

nouns that were neither very specific nor very general, and/or their extension was unclear, the initial classification of the combination was RC?.⁸

For example, *perform*:

OALD: to do sth, such as a piece of work, task or duty: *to perform an experiment/a miracle/a ceremony*

CCED: When you **perform** a task or action, especially a complicated one, you do it.

For these cases (i.e. in which it was unclear whether the sense of the verb is restricted or not), three nouns that were assumed to be neither extremely common nor very uncommon were chosen, which, according to the definition in the two dictionaries should combine with the verb in question. If one of these nouns could be shown to be blocked (rejection by two native speakers or non-occurrence in the British National Corpus⁹ and rejection by one native speaker were taken as evidence),¹⁰ the combination was classified as RC. If all of them were shown to be clearly acceptable (i.e. accepted by two native speakers or by one native speaker together with at least two occurrences in different texts in the British National Corpus), the combination was classified as F. If the evidence was inconclusive, the classification RC? was kept. Thus, the combinations the learners had produced were classified into F, RC?, RC, and I. This procedure has a few drawbacks as it relies on the distinction of senses and combinatory possibilities of headwords as they are presented in dictionaries and on a small sample of additional native speaker and corpus information. But it is believed that the distinctions made on its basis nevertheless approximate language reality as closely as possible in a study on the present scale.

As the classification into types of combination crucially depends on the verb, a special procedure had to be adopted for those combinations in which the verb was incorrect. Those combinations were re-classified after the native speaker acceptability judgements (which will be described in the next paragraph), and on the basis of the corrections provided by the informants. If the learner had apparently confused two verbs with unrelated senses, the combination was classified according to whether the intended sense of that verb is restricted or not; for example **remark (for notice) a car* was classified as F. If this was not the case, it had to be decided whether the verb was used in a sense that was closer to an unrestricted or a restricted sense of the verb used, again as evidenced by dictionaries. For example, **organize playgrounds* was classified as F because the sense of *organize* that the probably intended sense ('build') is closest to can be used freely with all nouns denoting events or activities (cf. *CCED*). On the other hand, **take a film for make/shoot a film* are classified as RC because the verb seems to be used as in *take a picture/ photograph*.¹¹ It hardly needs pointing out that this aspect of the classification procedure also remains slightly problematic, as it is subjective to some degree.

In the third step of the analysis, that is the judgements on the acceptability of the combinations produced by the learners, a variety of methods was

employed as well. The judgements take into account the combinations in their entirety (including, for example, the pre-modification and complementation pattern of the noun). Combinations were judged correct if they were found in the same form (including, for example, prepositions and articles) in the *OALD*, the *CCED*, *The BBI Dictionary of English Word Combinations*, the *Oxford Dictionary of Current Idiomatic English* (Benson *et al.* 1997), or in at least five different texts in the British National Corpus.¹² All the word combinations that could not be judged as correct on this basis were presented to two native speakers (one British, one American) with enough context to make the intended meaning clear (usually one sentence). The native speakers were asked to judge the word combination C (correct), W (wrong), or CW for 'not sure' or 'ok, but not the best way to say it', considering the context in their judgement. They were also asked to provide a correction if they judged the combination CW or W. If the judgements were C and W for the same combination or different elements of the combination were corrected, two additional native speakers were presented with the combination. If two native speakers of the same variety judged a combination C, it was assumed to be correct in that variety and therefore judged C. In all the other cases, an average of the judgements was worked out, which resulted in a five-stage scale of acceptability: clearly acceptable C, largely acceptable {C}, unclear CW, largely unacceptable {W}, clearly unacceptable W. Because of the variation in native speaker norms and because the scale is based on a comparatively low number of acceptability judgements for each combination, this scale, similar to the one for the degree of restriction, has to be regarded as an approximation.

The application to the data of the three steps outlined above thus resulted in a database of verb–noun combinations classified as to their degree of restriction (F–RC–RC?–I) and to their degree of acceptability (C–{C}–CW–{W}–W).

4. THE USE OF COLLOCATIONS BY ADVANCED LEARNERS

4.1 Types of mistakes

Altogether, 1072 verb–object–noun combinations were extracted from the learner essays, of which 213 were classified as collocations (RC or RC?), 846 as free combinations (F) and 13 as idioms (I). Table 1 shows their distribution on the scale of acceptability.

If those combinations judged either {W} or W by native speakers are regarded as 'wrong' or as containing 'mistakes', almost a quarter (255 out of 1072) of the combinations produced by the learners can be said to contain one or several mistakes;¹³ 56 of those are RCs. The mistakes are fairly evenly distributed over the essays (cf. Table 2).

As pointed out before, however, not only the verb and the noun of a combination were judged as to acceptability, but also the non-lexical elements belonging to a combination, so that, for example, *raise the question *about*

Table 1: Overall distribution (degree of restriction and acceptability)

Degree of acceptability	C	{C}	CW	{W}	W	Total
Free combinations (F)	579	46	25	61	135	846
Collocations (RC/RC?)	137	17	3	13	43	213
Idioms (I)	8	–	2	1	2	13
Total	724	63	30	75	180	1072

Table 2: Distribution of wrong RCs over essays

Number of wrong RCs	0	1	2	3	4	5	6	7
Number of essays	11	7	5	3	2	3	0	1

(BAS-0024.1;¹⁴ correct: *raise the question of*) was judged unacceptable. This means that by no means all mistakes occurring in RC-combinations are a mismatch between the verb and the noun. The types of mistakes were identified on the basis of the corrections the native speaker judges had given; if different elements of a combination were corrected, the element that was corrected by the majority of informants was considered the wrong one.¹⁵ The assignation of every mistake to a certain type is therefore a slight simplification but was deemed necessary for further analysis. The nine types of mistake that have been identified for the 56 unacceptable collocations are listed in Table 3.

In one case, the type of mistake could not be determined: *reach an aim* (DR-0001.1) was corrected to *reach a goal* by two of the judges and to *achieve an aim* by the other two. Owing to the fact that, in addition to this mistake (which is not included in Table 3), some combinations contained several mistakes (e.g. in *pass one's judgements* for *pass judgement* a pronoun and a number mistake were counted), the total number of mistakes in the 56 combinations was 65.

Of all these types of mistakes, the one occurring most frequently is the wrong choice of verb. This is not surprising since the verb in a collocation according to my definition has a restricted sense, which naturally makes its correct use difficult. Other examples are **make one's homework* (AUG-0001.3; *do*), **give a solution to* (BAS-0014.1; *provide*), **take one's task* (BAS-0014.1; *carry out/perform*). Also fairly frequent is the wrong choice of noun and the production of a completely wrong combination (called 'usage 2' in Table 3), followed by prepositional mistakes and determiner mistakes. Although mistakes in non-lexical elements (including prepositions) are less common, they still occur in more than a quarter of all wrong combinations, which indicates that it is not sufficient for the learner to know which lexical items collocate (such as *get + permission*, *fail + exam*), but that in order to produce an

Table 3: Types of mistakes in collocations

Type of mistake	Example	Occurrences
Verb	*carry out races	
Wrong choice of verb (or non-existent verb)	(AUG-0001.3; hold races)	24
Noun	*close lacks	
Wrong choice of noun (or non-existent noun)	(DR-0001.1; close gaps)	14
Usage 1	take notice	
Combination exists but is not used correctly	(AUG-0091.3; to notice)	2
Usage 2	*hold children within bounds	
Combination does not exist and cannot be corrected by exchanging single elements	(AUG-0004.3; ~show children where the boundaries lie)	11
Preposition (verb)	*fail in one's exams	
Preposition of a prepositional verb missing, present though unacceptable, or wrong	(AUG-0001.1; fail one's exams)	3
Preposition (noun)	*raise the question about	
Preposition of a noun missing, present though unacceptable, or wrong	(BAS-0024.1; raise the question of)	4
Determiner	*get the permission	
Article or pronoun missing, present though unacceptable, or wrong	(AUG-0008.2; get permission)	4
Number	*pass one's judgements	
Noun used in singular instead of plural or vice versa	(AUG-0075.3; pass judgement)	2
Structure	*make sb. friends	
Syntactic structure wrong	(AUG-0074.3; make friends with sb.)	1

Essay code and correction given in brackets in example column

acceptable combination, it is essential to know the whole combination (*get permission (to)*, *fail an exam*).

4.2 The role of the degree of restriction of a combination

To find out how the degree of restriction of a combination and the mistakes made by the learners are related, a further subdivision was added to the initial classification of F, RC?, RC, and I. The category RC was subdivided into collocations involving only a little restriction (called RC2) and collocations involving a lot of restriction (called RC1). This distinction was made on the

basis of whether a combination had been classified as collocation on the basis of criterion 2 alone (i.e. combinations in which the verb cannot, in the given sense, be used with every noun that would be syntactically and semantically possible; cf. section 2) or—additionally or exclusively—on the basis of criterion 1 (the verb can, in the given sense, only combine with a limited number of nouns). The former group (RC2) includes collocations such as *exert influence*, where possible nouns include *control, pressure, authority, power, attraction*, and quite a few others but not, for example, *rights*; the latter group (RC1) includes collocations such as *fail an exam/test*, where not many nouns are possible. On account of the fact that of those collocations classified RC? none had a verb restricted to a small set of nouns, and, in addition, since it is less certain that these combinations can be considered restricted at all, the following scale of restriction can be set up (from most to least restricted): I–RC1–RC2–RC?–F. The distribution of combinations and mistakes according to this refined classification is given in Table 4.

The table indicates that the highest rate of mistakes occurs in combinations with a medium degree of restriction.¹⁶ The lowest rate of mistakes, on the other hand, is found with combinations classified as RC1 (such as *pay attention* or *run a risk*). It therefore seems that whereas learners are mostly aware of the restriction in combinations where the verb only takes a few nouns, they are less aware of restrictions in combinations where the verb takes a wider range of nouns (such as *exert, perform, or reach*). An explanation for this phenomenon could be that combinations of the type RC1 are more often acquired and produced as wholes, whereas combinations of the type RC2 are more creatively—and sometimes too creatively—combined by learners.

In free combinations and idioms, only a few minor types of mistakes that were not also observed in the analysis of collocations were identified (e.g. wrong verb complementation such as **end up with commercials* (AUG-0093.1); instead of *end up watching commercials*). As can be seen in Table 5, the major types of mistakes that occurred were the same as in collocations (at least in free combinations; for idioms the numbers are too low for claims of this kind).

Although, if seen in relation to the whole number of combinations produced, verbs were not wrong as frequently in free combinations as in

Table 4: Number and percentage of mistakes (*{W}*, *W*) in the different types of combinations

	F	RC?	RC2	RC1	I	Total
Total	846	47	88	78	13	1072
Mistakes: n	196	13	29	14	3	255
Percentage	23%	28%	33%	18%	23%	24%

Table 5: The distribution of the major types of mistakes in free combinations and collocations

Types of mistake	F: number of mistakes	RC: number of mistakes	F: mistakes per combinations produced	RC: mistakes per combinations produced
Verb	65	24	7.6%	11.3%
Noun	59	14	7.0%	6.6%
Usage (1+2)	27	13	3.2%	6.1%
Preposition (verb+noun):	17	7	2.0%	3.3%
Determiner	22	4	2.6%	1.9%

collocations, they were also the major source of mistakes in free combinations—usually because a verb was confused with another one of similar meaning:

- cause/produce: . . . moving by means of two wheels, which I found much more exciting than the slow movement my feet were able to cause. (AUG-0056.3)
- evaluate/analyse: what mysterious methods you might have for evaluating your endless questionnaires (AUG-0004.4)

The relative frequency of the other major types of mistakes is also very similar for free combinations and collocations (see Table 5). It seems, therefore, that the degree of restriction does not have a major influence on the types and amount of mistakes learners make, except that collocations with a low degree of restriction are the most difficult kind of combination for the learners.

4.3 The role of the learners' L1 on collocation production

The role of the learners' first language was first investigated only with respect to wrong and questionable combinations (i.e. W, {W}, CW); later, possible first language influences on all combinations produced by the learners were examined. In the case of the wrong and questionable combinations, it was determined first whether the German equivalent of what the learners apparently attempted to produce was in any way (form and/or meaning) similar to what was actually produced. Since in a study such as the present one there is no way of ascertaining whether L1 influence actually occurred, similarity was considered an indication that influence was likely.¹⁷ For example, if **make homework* was produced by a learner, the fact that German has *Hausaufgaben machen* and that *machen* is related to *make* in both meaning and form led to the assumption of L1 influence.

In the first step of the analysis, then, I examined how many of the wrong or not quite acceptable combinations are likely to have been influenced by the

L1 and whether this influence was different according to the degree of restriction of a combination.

Table 6 shows that whereas the influence of the L1 on verb–noun combination mistakes is considerable in general (45 per cent), it is greatest in collocations (56 per cent). This result runs counter to the claims of some researchers on learner collocations who have hypothesized that L1 influence is not very important in collocational mistakes (e.g. Dechert and Lennon 1989; Lennon 1996). Assuming that the L1 influence on mistakes correlates with a high L1 influence on collocational production in general, this result supports claims of other researchers who (usually on the basis of much smaller amounts of data or even on the basis of impressions only) assume that first language influence on the production of collocations is rather high (e.g. Bahns 1993, 1997; Gabrys-Biskup 1990, 1992).

Table 6: L1 influence on mistakes and questionable combinations

	F	RC	I	Total:
W, {W}, or CW combinations	220	59	4	283
L1 influence likely: n	92	33	2	127
Percentage	42%	56%	50%	45%

In a second step, I investigated which types of mistakes were likely to have been influenced by L1 and how great this influence was. One important result of this analysis was that there was not a single type of mistake in which the L1 did not seem to play a role. Some examples of collocation mistakes that were probably influenced (though not usually exclusively caused) by the learner's first language are:

- verb mistake: **make homework* (BAS-0014.1; correct: *do homework*; German *Hausaufgaben machen*)
- noun mistake: **close lacks* (DR-0001.1; correct: *gaps*; German *Lücken schliessen*; phonological influence likely)
- usage mistake: *train one's muscles* (AUG-0092.1; correct: *to exercise*; German *seine Muskeln trainieren*)
- preposition mistake: **draw a picture from* (AUG-0019.3; correct: *of*; German: *ein Bild zeichnen von*; both *of* and *from* frequently correspond to German *von*)
- article mistake: **get the permission* (AUG-0008.2; correct: *get permission*; German *die Erlaubnis bekommen*)

A closer examination of L1 influence on the different types of collocation mistakes revealed that the L1 not only seems to have an influence on all types of mistakes, but that, in addition, the influence on all types of collocational mistakes is of remarkably similar strength. As shown in Table 7, in every one

Table 7: L1 influence on different types of mistakes in collocations

Type of mistake	Total	L1 influence likely
Verb	24	12
Noun	14	8
Usage	13	7
Preposition	7	4
Article	4	2

of the five groups of major types of mistakes, about half of the mistakes were probably influenced by the learner's L1.

Since L1 influence played such an important part in word combination mistakes, a type of analysis was carried out that to my knowledge has not been carried out before: an investigation of L1 influence on correct combinations as well as on the wrong and questionable ones. For that purpose, all combinations in the corpus were classified as to whether what was said (if correct) or what was probably intended (if wrong or questionable) is congruent in English and German. A very strict definition of congruence was applied, so that only combinations that sounded natural in both languages if they were rendered word for word were regarded as congruent (considering, however, general syntactic rules of the two languages). According to this definition, combinations such as *build a house—ein Haus bauen*, or *send a letter to him—ihm einen Brief schicken* are considered congruent, whereas combinations such as *make a decision—eine Entscheidung treffen*, *take care of—sich kümmern um* are considered non-congruent, as well as combinations such as *depend on his attitude—von seiner Haltung abhängen* (because of the difference in preposition). It came as no surprise that the general trend that emerged from this classification was that more restricted combinations were less often congruent than freer combinations. More interesting were the results obtained when the percentage of mistakes was compared in these two groups, that is the group of congruent and of incongruent combinations (cf. Table 8).

Table 8 clearly shows that the non-congruent combinations were consistently—that is independently of the degree of restriction of the combination—far more difficult for the learner than the congruent ones. In Table 9, this dimension congruence/non-congruence is compared to the dimension of restriction (F/RC/I) in relation to the degree of acceptability of the combinations produced. A correct/wrong ratio is given for each of these groups, with combinations that were judged W or {W} considered as 'wrong' and those judged C or {C} considered as 'correct' (CW combinations are not considered).

Table 9 shows that the ratio of correct versus wrong combinations is

Table 8: Mistakes (*{W}*, *W*) in congruent and non-congruent combinations

	F	RC?	RC2	RC1	I	Total
Congruent combinations	536	24	33	31	1	625
Mistakes: n	61	2	3	2	—	68
Percentage	11%	8%	9%	7%	—	11%
Non-congruent combinations	310	23	55	47	12	447
Mistakes: n	135	11	26	12	3	187
Percentage	44%	48%	49%	26%	25%	42%

Table 9: The dimensions congruence and degree of restriction and their relation to correctness

	Congruent	Non-congruent	F	RC	I
Total	625	447	846	213	13
C or {C}: n	544	243	625	154	8
Percentage	87.0%	54.4%	73.8%	72.3%	61.5%
{W} or W: n	68	187	196	56	3
Percentage	10.9%	41.8%	23.2%	26.3%	23.1%
Correct/wrong ratio	8:1	1.3:1	3.2:1	2.8:1	2.7:1

radically different for congruent combinations and non-congruent combinations (8:1 versus 1.3:1), whereas it is rather similar for free combinations (3.2:1), restricted combinations (2.8:1) and idioms (2.7:1). It is thus glaringly obvious that although the degree of restriction has some influence on whether the learner produces an acceptable combination or not, the fact of whether what the learner intends to say can be expressed in English in exactly the same way as it can be expressed in German has a far greater influence on the acceptability of what is produced. L1 influence in the area of word combinations therefore seems to be considerably stronger than even those researchers who have suspected its importance have assumed (cf. e.g. Gabrys-Biskup 1992). What implications this result and the others presented in this part of the paper have for language teaching will be discussed in the next section.

5. SOME IMPLICATIONS FOR TEACHING

As the present study has shown, even advanced learners have considerable difficulties in the production of collocations; the study has thus underlined what has recently been suggested by a number of authors, namely that

collocations do deserve a place in language teaching. In most of the recent publications on the teaching of collocations, however, making learners aware of the phenomenon is considered the foremost (and sometimes even only) task for the teacher. But whereas it is doubtless essential for learners to realize that a substantial number of those combinations that are easily understood can nevertheless cause problems in production because the elements cannot be combined freely, teaching collocations should not end once this insight has been provided. Although rote learning has fallen into discredit along with behaviourism, it seems indispensable that a number of collocations be taught and learnt explicitly. Since it is, however, clearly impossible to teach all (or even most) of the collocations in a language, criteria have to be set up to determine which collocations should be included in a given syllabus.

Obvious criteria for the selection of collocations to be taught are that they should be both undoubtedly acceptable and frequent in a neutral register and any special register that is of use to the learner (in an academic writing course, for example, combinations such as *conduct/do/carry out a study* or *make an analysis* would seem useful). As can be concluded from the present study, two additional criteria need to be taken into account when selecting collocations for teaching, namely congruence and restriction. The results strongly suggest that non-congruence in L1 and L2 is one of the most important criteria, that is, that collocations that are non-congruent should receive particular attention in language teaching. As the same appears to be true for free combinations, the tendency of the past few decades to downplay L1 influence and to disregard the L1 in foreign language teaching seems to be misguided. What cannot be inferred from the results, however, is, that congruent collocations can be entirely neglected because the learner will automatically get them right (as has been suggested by Bahns (e.g. 1993, 1997)). The analysis has shown that mistakes are also made when collocations are congruent; further research is needed, however, to find out which of the congruent collocations are difficult for the learner. With respect to restriction, it has been shown that less restricted collocations are particularly problematic for the learner, which means that at least in advanced courses special emphasis should be put on teaching RC2 combinations such as *exert pressure* or *perform a task*. The study has revealed, however, that degree of restriction is not as important a factor as might be expected and that in free combinations verbs are also very difficult for learners and should receive particular attention—again especially those that are non-congruent in L1 and L2.

On the basis of the material presented, some suggestions can also be made about how to teach collocations. First, it is not sufficient to merely teach the lexical elements that go together, but it is necessary to teach entire combinations including prepositions, articles, etc. (e.g. knowing that *pass* can combine with *judgement* is less useful, as could be seen, than knowing that it is *pass judgement on* and not **pass one's judgement* or **pass judgements* or **pass judgement about* or indeed anything else). Using grids, which is probably the most frequently encountered suggestion in the

literature on teaching collocations, is therefore not ideal for presenting and practising collocations.

Secondly, not only should the collocations be selected for teaching with reference to L1, but those collocations (or at least some of them) that have been selected on this basis should also actually be taught with reference to L1. Learners have to be made aware of L1–L2 differences, otherwise, despite having learnt the correct collocation, they are still likely to produce the L1 equivalent. It seems pointless, for example, to teach German-speaking learners the collocation *have an experience* without alerting them to the fact that **make an experience* (the equivalent to German *eine Erfahrung machen*) is not possible in English. Contrasting L2 and L1 collocations should, however, not be limited to lexical elements either, but should include contrasts in the use of non-lexical elements such as articles and prepositions (for example *get permission* vs. *die Erlaubnis bekommen*). As the results also reveal that the difficulties learners have with free combinations are similar to those they have with collocations, teaching free combinations in a way similar to collocations seems desirable as well, in particular pointing out what (groups of) nouns are possible with certain verbs. And as the role of congruence also seems to be of considerable importance for the production of free combinations, contrasting L1 and L2 can also be recommended for free combinations.

A third suggestion on how to teach collocations that can be derived from the present study is that in the teaching of verb–noun collocations the focus should be on the verb, since it is the verb that causes the greatest difficulties. However, depending on whether one deals with collocations of type RC1 or RC2, different approaches need to be adopted. In collocations of the RC1 type, it is sufficient to teach the one or two (or at most around five) combinations that the verb is used in and to alert the learner to the fact that the verb in that sense is limited to these combinations. For example, if the collocation *run the risk (of)* is taught, it should be pointed out that other combinations such as **run the danger (of)* or **run the peril (of)* are not possible.¹⁸ RC2 collocations, on the other hand, call for a different procedure. The most important point when dealing with this type of collocation is to make the learner aware that the verb cannot be used freely. Of the possible combinations, then, only the frequent and acceptable ones as well as the non-congruent ones should be taught, and some semantically possible but collocationally impossible combinations with the verb in question should be highlighted, especially those that are possible in the learners' L1. For the verb *reach* in the sense of 'succeeding in achieving something', for example, the combinations *reach a decision/a conclusion/a compromise/an agreement/a goal* could be taught, and it should be pointed out that although the verb can be used widely in this sense, it cannot be used completely freely, so that, unlike the corresponding collocation in German, **reach an aim* (German: *ein Ziel erreichen*) is not acceptable in English. Contrasting the verb and its uses with others that have a similar sense (such as *come to* and *achieve* in this case) would also seem a helpful procedure for RC2 collocations because it highlights the differences in the restriction of

these verbs (e.g. the existence of *achieve an aim* as opposed to **reach an aim* and **come to an aim*).

Obviously, the suggestions made in this section will not only have to be confirmed by further learner language analysis (including research on other grammatical collocation types and on other language pairs), but will also have to be tested as to their pedagogical effectiveness. In addition, many of the suggestions can only be realized in groups of learners with the same native language and if the teacher knows the native language of the learners at least to a certain degree. For other groups, focusing on frequent, acceptable and less restricted combinations, or more precisely, on their verbs, seems the most reasonable approach to teaching collocations at this point. In those many cases where it is possible, however, an L1-based approach to the teaching of collocations seems highly desirable.

(Revised version received October 2002)

ACKNOWLEDGEMENTS

I would like to thank the coordinator of German ICLE, Gunter Lorenz, and Sylviane Granger and the Centre for English Corpus Linguistics at the Université Catholique de Louvain, Belgium, for integrating me into the ICLE project at a late stage, and the students from the University of Basel, Switzerland, who contributed essays to my collection for ICLE. Thanks also go to all native speakers who provided me with acceptability judgements, most especially to Peter Burleigh at the University of Basel. I would also like to thank David Allerton and Christian Mair as well as the three anonymous reviewers for their insightful comments on previous versions of this paper.

NOTES

- 1 The study is part of a larger study, in which eventually all the essays (around 320) making up the German subcorpus of ICLE (International Corpus of Learner English) will be examined with respect to collocation use.
- 2 *Oxford Advanced Learner's Dictionary* (2000).
- 3 *Collins COBUILD English Dictionary* (1995).
- 4 The ICLE-project is coordinated at the Centre for English Corpus Linguistics at the Université Catholique de Louvain, Belgium, cf. <http://www.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/introduction.html>.
- 5 The following criteria were applied: 20 essays each were selected from the batches of essays collected in Augsburg, 10 from the batch collected in Basel and 2 from the batch collected in Dresden. If an essay selected was written by a student with an English-speaking parent or was on a topic that was already covered by another essay that had been selected, this essay was replaced by another one.
- 6 Verb-object-noun combinations 'hidden' in transformed constructions (such as passive or relative clauses) were also included. If one verb was combined with several nouns (as for example *arouses strong contrary feelings and opinions*; DR-0001.1), several combinations were assumed to be present (in this example *arouse feelings* and *arouse opinions*).
- 7 In the *CCED*, a conscious attempt is made to include the combinatory possibilities in the definitions (1995: xviii). The *OALD*, though nowhere explicit on this point, follows the same practice and, in addition, uses bold print, slashes and 'etc.' to indicate possible restrictions on the combinability of head-words.
- 8 The final classification of *perform* according to the procedures described below was RC.

- 9 The British National Corpus consists of 100 million words of modern British English. It contains about 90 million words of written language and 10 million words of spoken language of various text types ranging from formal academic prose and popular fiction to transcribed radio shows and informal conversation (cf. <http://www.hcu.ox.ac.uk/BNC>).
- 10 All native speakers who participated in this study are native speakers of either British or American English, and either hold a university degree or are university/college students from their third year onwards.
- 11 The same procedure was followed with respect to wrong idioms: if a noun did not seem to have an unrestricted meaning, that is an independent meaning outside of the combination, the combination was classified as an idiom. This was only the case once: *see sb's spearhead* (AUG-0091.3) was apparently used in roughly the sense 'realize that sb. is ready to fight for their rights'.
- 12 The threshold chosen is rather high because it emerged in the course of the analysis that native speaker intuition on the acceptability of word combinations is usually stricter than native speaker production.
- 13 'Mistake' and 'error' are not differentiated in this paper, and neither are 'unacceptable' and 'wrong'.
- 14 The essay codes are given in a slightly simplified form in this paper: their full form reads ICLE-GE-BAS-0024.1, but 'ICLE-GE-' is omitted throughout this paper.
- 15 As described in section 3, four native speakers were asked for judgements in such a case; sometimes single informants provided more than one correction, which were then all taken into account in order to determine the type of mistake. If no correction was provided by one of the informants, only the one(s) provided were considered.
- 16 As this is an exploratory study with relatively little data, no statistical tests were carried out.
- 17 Although a few of the learners whose data are included in German ICLE have a good knowledge of other languages in addition to German and English, German is the strongest language of all the learners and therefore probably most likely to affect L2 production. It cannot be excluded, however, that a few of the assumed L1 influences are in fact L3 influences or a combination of L1 and L3 influences.
- 18 Although *run the danger* is sometimes used, it is considerably less common than *run the risk*, usually judged wrong by native speakers (all four native speakers who were asked to judge the acceptability of *run the danger* considered it wrong or of doubtful acceptability), and therefore should not be taught. (In the British National Corpus, there are 11 occurrences of *run the danger* in 10 texts, whereas there are about 330 occurrences of *run the risk* in about 310 texts; the span considered was +/-6.)

REFERENCES

- Bahns, J. 1993. 'Kollokation kontra Kontext: Wider ein zu weites Verständnis des Kollokationsbegriffs.' *Praxis des neusprachlichen Unterrichts* 40: 30–7.
- Bahns, J. 1997. *Kollokationen und Wortschatzarbeit im Englischunterricht*. Tübingen: Narr.
- Bahns, J. and M. Eldaw. 1993. 'Should we teach EFL students collocations?' *System* 21/1: 101–14.
- Benson, M., E. Benson, and R. Ilson. 1997. *The BBI Dictionary of English Word Combinations*. (Rev. ed. of *BBI Combinatory Dictionary of English*, 1986.) Amsterdam: Benjamins.
- Chi, M. L. A., P. K. Wong, and C. M. Wong. 1994. 'Collocational problems amongst ESL learners: a corpus-based study' in L. Flowerdew and A. K. Tong (eds): *Entering Text*. Hong Kong: University of Science and Technology, pp. 157–65.
- Collins Cobuild English Dictionary. 1995. London: Harper Collins.
- Cowie, A. P. 1992. 'Multiword lexical units and communicative language teaching' in P. J. L. Arnaud and H. Béjoint (eds): *Vocabulary and Applied Linguistics*. Houndmills: Macmillan, pp. 1–12.

- Cowie, A. P. 1994. 'Phraseology' in R. E. Asher (ed.): *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon, pp. 3168–71.
- Cowie, A. P., R. Mackin, and I. R. McCaig. 1993. *Oxford Dictionary of English Idioms*. Oxford: Oxford University Press.
- Dechert, H. W. and P. Lennon. 1989. 'Collocational blends of advanced second language learners. A preliminary analysis' in Wieslaw Olesky (ed.): *Contrastive Pragmatics*. Amsterdam: Benjamins, pp. 131–68.
- Farghal, M. and H. Obiedat. 1995. 'Collocations: a neglected variable in EFL.' *IRAL* 33: 315–31.
- Gabrys-Biskup, D. 1990. 'Some remarks on combinability: lexical collocations' in J. Arabski (ed.): *Foreign Language Acquisition Papers*. Katowice: Uniwersytet Slaski, pp. 31–44.
- Gabrys-Biskup, D. 1992. 'L1 influence on learners' renderings of English collocations. A Polish/German empirical study' in P. J. L. Arnaud and H. Béjoint (eds): *Vocabulary and Applied Linguistics*. London: Macmillan, pp. 85–93.
- Granger, S. 1998. 'Prefabricated patterns in advanced EFL writing: collocations and formulae' in A. P. Cowie (ed.): *Phraseology. Theory, analysis, and applications*. Oxford: Clarendon Press, pp. 145–60.
- Hausmann, F. J. 1989. 'Le dictionnaire de collocations' in F. J. Hausmann, H. E. Wiegand, and L. Zgusta (eds): *Wörterbücher, dictionnaires, dictionnaires. Ein internationales Handbuch zur Lexikographie*. Berlin: de Gruyter, pp. 1010–19.
- Howarth, P. 1996. *Phraseology in English Academic Writing. Some implications for language learning and dictionary making*. Tübingen: Niemeyer.
- Hussein, R. F. 1990. 'Collocations: The missing link in vocabulary acquisition amongst EFL learners.' *Papers and Studies in Contrastive Linguistics* 26: 123–36.
- Kennedy, G. D. 1990. 'Collocations: Where grammar and vocabulary teaching meet' in S. Anivan (ed.): *Language Teaching Methodology for the Nineties*. Singapore: SEAMEO Regional Language Centre, pp. 215–29.
- Lennon, P. 1996. 'Getting "easy" verbs wrong at the advanced level.' *IRAL* 34: 23–36.
- Lewis, M. (ed.) 2000. *Teaching Collocation. Further developments in the lexical approach*. Hove: LTP.
- Lorenz, G. R. 1999. *Adjective Intensification—Learners Versus Native Speakers. A corpus study of argumentative writing*. Amsterdam: Rodopi.
- Mel'čuk, I. 1998. 'Collocations and lexical functions' in A. P. Cowie (ed.): *Phraseology. Theory, Analysis, and Applications*. Oxford: Clarendon Press, pp. 23–53.
- Oxford Advanced Learner's Dictionary*. 2000. Oxford: Oxford University Press.
- Sinclair, John McH. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stubbs, M. 2001. 'Texts, corpora, and problems of interpretation: A response to Widdowson.' *Applied Linguistics* 22/2: 149–72.
- Wray, A. 2002. *Formulaic Language and the Lexicon*. New York: Oxford University Press.