Transfer and universality: Collocation use in advanced Chinese and Swedish learner English

Ying Wang, Uppsala University
Philip Shaw, Stockholm University

Abstract
Research hitherto has tended to show that, like vocabulary errors, wrong collocations are often due to transfer from L1, and hence that greater typological difference from the target language leads to more serious collocational errors.
We tested this hypothesis by comparing the collocational errors of Swedish and Chinese students. 100 university students from each environment wrote a short essay following the same prompt. The most common verbs were identified and then all occurrences of verb + noun collocations with have, do, take *and* make *were investigated. Collocations were divided into free and restricted and then into target-like and non-target-like. The results showed that the two groups chose similar sets of noun collocates and made similar types and proportions of errors. This suggests that intralingual factors are as important as L1 transfer in this area.*

1   Introduction
1.1   Collocations: A neglected variable in language teaching and learning
It has been widely accepted that collocations, i.e. recurrent word combinations such as *break the silence*, *deeply absorbed*, are fundamental to all language use (Wallace 1984; Partington 1998; Erman and Warren 2000; Singleton 2000; Hill 2001; Stubbs 2001; Grant and Bauer 2004). Bolinger (1976: 14) was one of the first to point out that our language does not expect us to "build everything starting with lumber, nails, and blueprint". Instead, it provides us with an incredibly large number of conventionalized multi-word combinations. Support for this view has been provided by research in corpus linguistics over the last few decades (e.g. Sinclair 1991; Altenberg 1998; Stubbs 2001). Through looking into large collections of naturally-occurring texts, corpus studies challenge the traditional view that knowing a language involves two types of knowledge:

grammatical rules and individual lexical items. Further evidence has come from neurophysiological and psychological studies, which indicate that the human mind is better equipped for memorizing than for creative processing. The use of ready-made multi-word expressions reduces the processing effort and thus plays a major role in language production and comprehension (Pawley and Syder 1983; Cantos and Sánchez 2001; Wiktorsson 2003; Nesselhauf 2005).

When it comes to language learning, a range of arguments have been put forward to justify giving attention to collocations. In learning another language, it is evident that we have to learn both grammatical correctness and idiomatic preference (Wiktorsson 2003). Collocations, as shown by corpus studies, constitute an important part of idiomaticity (e.g. Nesselhauf 2005). Pawly and Syder (1983) argue that collocational knowledge, as the essence of language knowledge, is indispensable for language learners to produce fluent and appropriate language. Many words are used in a limited set of collocations / multi-word units, and thus knowing their collocational possibilities should be one essential aspect of language learning.

However, given the secondary status that vocabulary was relegated to until the 1980s and 1990s in favor of emphasis on syntactic structures in the tradition of language teaching (Judd 1978; Carter and McCarthy 1988; Henriksen 1999; Bogaards 2001), it is not surprising that vocabulary has often been considered only as a reading or receptive problem. Since collocations usually do not pose any serious problems to language learners in the understanding process (Grant and Bauer 2004), the acquisition of collocations has been regarded as relatively unimportant. More precisely, the emphasis has been traditionally laid on improvement in quantitative rather than qualitative terms (Laufer and Nation 1995; Lennon 1996; Laufer 1998; Bogaards 2000, 2001) when vocabulary instruction is concerned. Language learners, who normally assume that word power lies in greater vocabulary size, rely heavily on single words as units of language production and often fail to recognize or respond to constraints of collocation for even the most well-known words (Levenston and Blum 1977; Morgan 2001). As a result, as Allerton (1984:39) observed, "so often the patient language-learner is told by the native speaker that a particular sentence is perfectly good English…, but that native speakers would never use it".[1]

The difficulties that language learners are faced with and the necessity for them to acquire collocational knowledge are now being gradually recognized and paid growing attention (e.g. Levenston and Blum 1977; Pawly and Syder 1983; Biskup 1992; Bogaards 2000; Aitchison 2003; Nesselhauf 2003, 2005). However, there still remains a lack of detailed description of learners' collocational performance as the basis for understanding how collocational competence

develops (Howarth 1998) and what learners' difficulties in this connection are, despite a handful of studies already conducted.

One type of collocational difficulty for learners, especially advanced learners, is associated with frequent, high-utility dynamic verbs. Ringbom (1998), for example, found that the use of *get* illustrates that advanced learners with 7 different L1s shared a lack of collocational competence. These highly frequent delexicalised verbs have been the focus of several studies highlighting learners' difficulties with collocations (e.g. Lennon 1996; Källkvist 1999; Altenberg and Granger 2001; Liu and Shaw 2001). Nesselhauf (2005) conducted one of the most systematic investigations to date of verb-noun collocations used by German learners of English. She found that most of the collocations that pose problems for the learners are frequent, everyday expressions. While these high frequency verbs are perfectly easy to understand in reading comprehension and are presumed to be acquired by language learners at an elementary or intermediate stage, the process of deducing usage restrictions, permitting native-like use in production in fact eludes many highly advanced learners (Källkvist 1999).

Language transfer, that is, the influence arising from a learner's conscious or unconscious judgment that something in the L1 is similar to the target language (TL), has important effects on second language acquisition (e.g. Blum and Levenston 1978; Ringbom 1983; Odlin 1989; Biskup 1992; Bahns 1993; Wang 2003; Jiang 2004; Chan 2004). Some studies have attempted to establish that L1 transfer is the main cause of the collocational problems observed. Bahns (1993), for example, through studying the learners whose mother tongue is closely related to English, argued that the learners' L1 may lead to appropriate as well as inappropriate collocational usages. It is well known that the distance between the languages concerned is one important determining factor in language transfer (Ringbom 1983; Odlin 1989; Chan 2004). Biskup (1992), in an attempt to ascertain whether the typological distance would have an influence on English learners' collocational use, observed learners of English whose L1s are German and Polish respectively, and found that L1 transfer is more likely in the case of two closely related languages. But one of the features shared by the research done by Biskup, Lennon, Bahns, and some other writers such as Farghal and Obiedat (1995), is that all these studies involved only translation tasks or word-association tests, and mostly relied on a very limited number of learners, making it questionable whether the results are generalizable, and whether they could reflect the learners' actual production problems (Nesselhauf 2003). Quite a few studies in recent years have analyzed language learners' collocational performance on the basis of their natural production (e.g. Granger 1998b; Howarth 1998; Ringbom 1998; Altenberg and Granger 2001; Liu and Shaw

2001; Nesselhauf 2003; Wiktorsson 2003) and made use of large samples of data as well. But almost all of them focused on one or another group of non-native learners, mostly European, apart from Biskup (1992), Ringbom (1998) and Altenberg and Granger (2001), who included learners with different (though related) L1s. More research is necessary to provide new insights into the relation between L1 transfer and ES/FL learners' collocational use. It would obviously be interesting, as suggested by Källkvist (1999), to explore this issue further by involving the interlanguage (IL) of learners whose L1 is not Indo-European in origin, e.g. Chinese, Japanese and Korean.

1.2   Definition and classification of collocation

Although the concept has long been a popular topic in linguistics, there is no universally accepted formal definition of collocation (Lewis 2001; Grant and Bauer 2004), which results in a plethora of different terms such as fixed expression, word-combination, idiom, phrase, and prefabricated pattern. However, we can still trace the term back to the 1950s, when it was first coined in its modern linguistic sense by the British linguist J. R. Firth (1957), who argued that "you shall judge a word by the company it keeps".

Some collocations, as argued by Partington (1996), are rather fixed, whereas some allow degrees of variation. Stated in another way, there is a scale of collocational restriction among different categories of collocations. At one end of the scale are the most restricted ones in which the occurrence of one item is conditioned by the other, whereas at the opposite end are unrestricted ones, which means one particular lexical item is open to partnership with a wide range of items. Accordingly, many linguists and researchers tend to classify multi-word units into the following three categories (e.g. Howarth 1998; Lewis 2001; Cowie 1993; Bahns 1993; Nesselhauf 2003; Grant and Bauer 2004), though the terminology used by them might be slightly different:

Free collocations (also referred to as open collocations or free word combinations) consist of items used in their literal senses and freely substitutable, such as *open the gate*, *a nice car*. This category seems to include all possible and semantically natural combinations.

Notice that saying a collocation is a free one does not mean that there is no restriction at all. The major difference between free collocations and restricted ones is that the restriction for the former is a result of the semantic properties of the two components concerned, whereas the restriction for the latter is "a somewhat arbitrary convention of the language" (Nesselhauf 2003: 225).

Restricted collocations (also referred to as fixed combinations or collocations) usually have one item used in a non-literal sense, often a specialized, or figurative sense, and the other used in its normal meaning such as *run a company, bitterly contested*. A collocation of this category, according to Howarth (1998), might permit limited substitution in either of its constituents as in *make/reach a decision* and *take on an obligation/a duty*, or in both components as in *do/carry out research/ a project*. The vocabulary choice is less predictable in this category of collocations than in the previous one.

Idioms are relatively frozen expressions, the meanings of which can barely be derived from the meanings of their constituent parts such as *sweeten the pill*, *kick the bucket*.

On another dimension, collocations can be divided into two major types depending on the word class of their constituents (Biskup 1992; Lewis 2001):

Lexical collocations combine two open class words such as verb + noun (*lead a life*), adjective + noun (*a vague answer*).

Grammatical collocations combine an open class word and one closed class word (grammatical word) such as preposition + noun (*in advance*), verb + preposition (*engage in*), or a grammatical structure such as an infinitive (*opportunity to do something*), clause (*to be afraid that…*).

1.3   Notion of collocational errors

Many researchers have attempted to analyze collocational errors made by EFL learners in order to demonstrate their difficulties in collocation use. A common problem that EFL learners have with collocation use, as observed by Lennon (1996) and others, is that they lack knowledge as to the collocational possibilities of verbs; hence mismatches between lexical items as in *stop the fire* (*put out the fire*). A second type of error is blending (Howarth 1998), i.e. to fill in the combinations within overlapping clusters by analogy, hence the wrong use of *pay care* (blend of *pay attention* and *take care*). So far, much attention has been focused on the collocational possibilities of the two lexical items in question. However, in a study of verb + noun collocations used by a group of German learners of English, Nesselhauf (2003) found that by no means all errors occurring are a mismatch between the verb and the noun. Other types of errors such as prepositional errors as in *raise the question about* (*raise the question of*) and determiner errors as in *get the permission* (*get permission*) are also fairly frequent among advanced EFL learners. These types of errors are particularly

related to restricted collocations due to their variability. According to Erman and Warren (2000: 52), many collocations allow inflectional variability, which involves the choice of tense, aspect, voice and a determiner (*to his/her/our surprise, lay a/the table*). However, the variability is not always predictable, but restricted in what seems to be a "non-generalizable manner", which accounts for the above types of collocational errors. Therefore it is not sufficient for EFL learners to know whether lexical items collocate (such as *make + decision*, *have + responsibility*). In order to produce acceptable language use, it is essential to know the whole combinations (*make a decision*, *have responsibility for doing something*).

1.4   Aims

As noted above, there is a need to investigate whether there are any differences in the collocational knowledge of advanced learners of English with different L1 backgrounds, and what problems they might have, shared as well as language-specific. One might hope to identify cross-linguistic principles besides L1 transfer which cause the problems. The present study compares the collocational features of a few high frequency verbs in English essays written by speakers of two languages, one of them (Swedish) close typologically to English, the other (Chinese), very remote. More specifically, following up Granger's suggestion (1998a: 13) that we can "gain a better insight into the nature of interlanguage" by comparing two learner-language or IL corpora, we attempt to make an empirical comparison of verb + noun collocations occurring in corpora made up of 200 elicited English essays written by the two groups of learners concerned. Furthermore, since restricted collocations are said to be an area where learners have considerable difficulties but which remains neglected (Howarth 1998), all the collocations are classified and later evaluated in terms of acceptability in an attempt to find out whether restricted collocations are indeed problematic in both learner corpora and whether the two groups' difficulties in this respect are similar or different.

2   Subjects and corpora

Two learner English corpora, or essays by Chinese-speaking learners of English (CSLE) and Swedish-speaking learners of English (SSLE) respectively, were built up for the purpose of comparison in the present study. They are of similar size, each consisting of 100 essays of around 200 words, which were written by the two groups of subjects with a 30-minute time limit in class without preparation in advance or access to any language tool when performing the task.

Both sets of subjects, despite their widely different learning environments, were at an advanced level of English proficiency in terms of extra-linguistic features such as the teaching level and the number of years for which the learners had learnt English. The subjects of CSLE were mostly third-semester English majors in the Foreign Languages School of Wuhan University, China. They were between 19 and 22 years old, and had learned English for about eight years in an input-poor environment for EFL learners. The subjects of SSLE were first-semester students in the English Department of Stockholm University, Sweden, between the age of 20 and 27, and had learned English for about nine years, but in an environment with a good deal of informal input from mass media.

The data is of the elicited kind, with the same topic as prompt: *Is it true that only rich countries can afford to worry about the environment?* One of the most important recent findings of IL study is that learners' IL varies across different situations and tasks (Källkvist 1999; Jarvis, *et al*. 2003). According to Granger (1998a: 8), the topic of a task is a relevant factor which may lead to the variation because "it affects lexical choice". Especially if learners are from different L1 and cultural backgrounds, as confirmed by Reid's (1990) study, the lexical choices in their essays will vary even more significantly with topic type. Since the focus of the present study is to compare the use of lexical collocations in the learners' IL, it would be more promising if the learners' lexical choice were not very irregularly dispersed in the first place. Therefore the same topic was utilized to stimulate all the subjects.

3   Methods

After all the handwritten essays had been typed out in an electronic format, a frequency wordlist for each corpus was generated by WordSmith Tools (Scott 1996). The result for the selected verbs were then further lemmatized manually so as to get the total occurrences of each word.

The next step was to generate concordance lines with each selected verb (in each of its inflected word-forms) as the key word, with the help of the KWIC (Key Word in Context) index in WordSmith Tools. It is strongly suggested by Collier (1993) and Sinclair (1991) that beyond four words from the key word there are no statistical indications of the attractive power of the key word. In the present study, we used a ± 5 span in order to take account of the cases in which there are pre-modification patterns (including articles, adverbs and adjectives, etc.) of the collocating nouns, which may occur either after or before the key-word verb.

Next we extracted manually all verb + noun collocations from the concordances we had got in the previous step. Then we attempted to classify them into the three categories, namely free collocations, restricted collocations and idioms. Inspired by Nesselhauf's (2003) method of determining whether a word is used in a restricted sense or not, we decided to use *The BBI dictionary of English word combinations* (BBI) as the main reference, supplemented by the *Collins Cobuild English dictionary* (CCED) because the former provides very detailed information on restricted collocations of English and the latter gives general combination information about a word. Since the few key verbs under scrutiny are high frequency words, which, according to Altenberg and Granger (2001), are characterized by a high degree of polysemy caused by their general and delexicalized uses, it is impossible for the dictionaries to give us a full picture of what nouns those verbs can combine with. So we chose to look up the noun in a verb + noun collocation instead, and if there was an indication of specific verbs for the noun in relation to the specific meaning the learner intended to express by the combination in either dictionary, the verb + noun collocation was considered as a restricted one. For example, we found *have a responsibility* and *do things* in the learner corpora. In the BBI, we found that for the same meaning of having obligation to do something, there is a set of verbs including *bear, shoulder, take, exercise, take on, have* which can be combined with the noun *responsibility*. Thus *have a responsibility* was considered a restricted collocation. Though it is not covered by the present study because we started from the verbs, we did find in the learner corpora the phrase *own a responsibility*, which obviously has the same meaning as *have a responsibility*, but *have* cannot be replaced by *own* here. This might serve as further evidence that such a collocation is a restricted one because it is clearly possible for a learner to use a word which is semantically acceptable but which native speakers would never use. For the same reason, a collocation like *do mistakes* was considered as a restricted but problematic one because in this case, the noun *mistakes* is supposed to have the verb collocate *make*. The same principle applied to all the collocations investigated in the present study. As for *do things*, there are no specific verbs indicated in either dictionary to go with *things*, but the combination is plausible by the rules of English syntax and semantics. That is to say, *do*, used as a transitive verb here, syntactically requires the presence of a Direct Object in the form of a Noun Phrase. Semantically, the noun *things* can be regarded as a general term for 'actions, activities', and thus is compatible with the verb *do*, which means 'to perform an action or activity', whereas some other nouns, *water*, for example, cannot be combined with *do*, simply because they are semantically incompatible. So the collocation *do things* was considered as a free

collocation in the study. As for idioms, since they are frozen expressions, the meaning of which is opaque from their constituent parts, it is much easier to recognize them.

Finally, the acceptability of the collocations was evaluated. All the collocations concerned, including the determiners, adjectives, adverbs, and prepositions they may take, were judged as correct or wrong/odd. In some cases, the context in which the collocations were used was taken into consideration as well. They were judged correct if they were found in the BBI, the CCED, or in at least five different texts in the British National Corpus (BNC). However, it is impossible for either dictionary to give a full picture of the collocational information, especially in terms of free collocations. The BNC has its drawbacks in use too; namely, it is difficult to extract inflectional variations or changes of the order of certain constituents, changes of determiner, or insertion of modifying items, etc. in searching for a collocation. To solve the problem in a less time-consuming yet more efficient way, some collocations, which we failed to find in the above sources, were presented to the second author as a native speaker with their relative concordance lines to make the intended meaning clear. The native speaker was asked to judge the collocations as acceptable, unacceptable/odd. If one collocation was judged as unacceptable or odd, the native speaker was also asked to point out the possibly problematic constituent(s) of the collocations. If he was not sure whether it was acceptable or not, three smaller English corpora (BROWN, FLOB, FROWN) were then employed together with WordSmith Tools, which allow the variations that the BNC does not in search of a sequence of words. Considering the size of the three corpora, a collocation was considered acceptable in the present study if it was found in the corpora, regardless of the number of its occurrences.

While this way of classification and evaluation turned out to be clear-cut and effective in most cases, there were some hard cases as well. Specifically, when the collocations were produced or misused by the learners, it is very difficult to say which category, namely free ones or restricted ones, they belong to. For instance, collocations such as *make factories*, *do recycling*, occurred quite frequently in the corpora. As for the former one, if the learner meant to express *operate a factory*, then according to our criteria, it should be regarded as a restricted one because there are a limited set of verbs such as *manage, run, operate* to combine with the noun in this sense. However, the learner seemed to mean '*build factories*', in which case there is no indication of specific verbs in either dictionary to combine with *factory*. Thus the two words were regarded as being freely combined. Similarly, there is no indication of specific verbs for *recycling* in the sense the learner meant to express either. It was regarded as a free colloca-

tion because the intended meaning is '*recycle paper/glass*', etc., in which both elements are used in their literal senses. Therefore in deciding which category a problematic collocation of this type belongs to, we also took its intended meaning into consideration. Since it is widely acknowledged that the line between free collocations and restricted ones is difficult to draw and by no means rigid, arbitrary decisions had to be made sometimes to enable investigation.

4   Results

4.1   The shared high frequency verbs used in both learner corpora and the occurrences of their verb + noun collocations

The first 20 verbs in the raw wordlists of both corpora (see Appendix I), were extremely similar, 16 of them being shared by both. After lemmatization, the 16 word-forms shrank to 11 words: *be, can, have, worry, afford, should, do, think, will, take* and *make*, among which *have* (excluding auxiliary verb usage), *do* (excluding auxiliary verb usage), *take* (excluding structures like *take it seriously*) and *make* (excluding structures like *make it possible*), were chosen to be the focus of the present study.

All the necessary word-forms related to the four key verbs were taken out and counted separately. Table 1 shows the total occurrences of each word, and of the verb + noun collocations under investigation. Figure 1 demonstrates more clearly that the frequencies of the verb + noun collocations for each key verb are similar in both corpora after normalization, though the SSLE employed slightly more verb + noun collocations on the whole.

Table 1:   Overall information about the two learner corpora

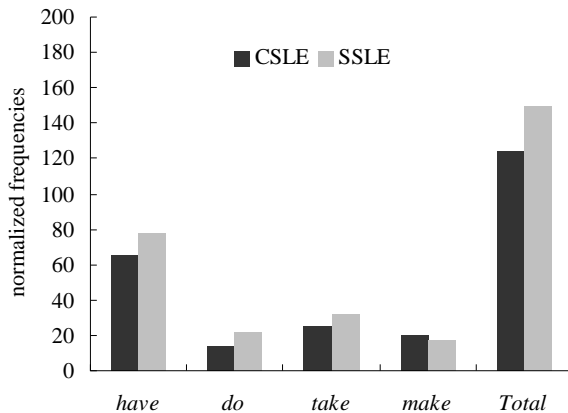|	CSLE	SSLE
Total tokens	24,102	19,995
No. of essays	100	100
No. of writers	100	100
Occurrence of the word *have* (*have* + noun collocations)	293 (156)	340 (156)
Occurrence of the word *do* (*do* + noun collocations)	150 (33)	205 (44)
Occurrence of the word *take* (*take* + noun collocations)	72 (61)	75 (64)
Occurrence of the word *make* (*make* + noun collocations)	74 (49)	76 (34)
Total occurrence of the above verbs (verbs +noun collocations)	589 (299)	696 (298)

Figure 1: Normalized frequencies (per 10,000 words) of distribution of the verbs + noun collocations investigated in the present study

Altogether, 597 verb + noun collocations were extracted from the learner corpora, of which 222 were classified as free collocations, 374 as restricted collocations and 1 as part of an idiom (*one can't make an omelette without breaking eggs*). Since there was only one occurrence of something that might be an idiom in the corpora, no further investigation about the use of idioms was carried out in the study.

4.2   *Lexical variety of the noun collocates for each verb*

Lexical variation or the richness of the noun collocates for each verb was measured by the type-token ratio (t/t), the ratio of the number of different nouns to the total number of collocations for each verb. A high t/t ratio in principle means that there is little repetition of lexical words, hence the wider lexical variation. Table 2 presents a general picture of the lexical variety of the noun collocates for each verb in the learner corpora. In general, the t/t ratio is slightly higher for the SSLE than for the CSLE. In terms of restricted collocations, however, both corpora demonstrated a very similar t/t ratio of the noun collocates for each verb, and the same ratio (37 per cent) in aggregate. When it comes to free collocations, the differences are relatively bigger than those of the restricted collocations. The SSLE corpus displayed greater variety in this relation, which in turn led to its slightly greater variety on the whole.

Table 2: Overall distribution of *have/do/take/make* + noun collocations in the learner corpora

		CSLE			SSLE		
		Token	Type	t/t (%)	Token	Type	t/t (%)
have + noun	Free	81	28	35	79	33	42
	Restricted	75	32	43	77	30	39
	All	156	60	39	156	63	40
do + noun	Free	12	2	17	21	6	29
	Restricted	21	9	43	23	13	57
	All	33	11	33	44	19	43
take + noun	Free	10	8	80	8	4	50
	Restricted	51	13	25	56	13	23
	All	61	21	34	64	17	27
make + noun	Free	9	2	22	2	2	100
	Restricted	40	16	40	31	13	42
	All	49	18	37	33	16	48
verbs + noun	Free	112	40	36	110	45	41
	Restricted	187	70	37	187	69	37
	All	299	110	37	297	114	38

4.3 Shared and exclusive noun collocates for each verb

Lists of the noun collocates for the four key verbs occurring in both corpora are given in Appendix II. Some of the noun collocates were shared, in that they occurred in both corpora, and some were used exclusively by either the Chinese-speaking learners or the Swedish-speaking learners. The distribution of shared and exclusive noun collocates occurring in the corpora is presented in Table 3, which, again, demonstrates the striking similarity between the two corpora. Specifically, more than half of the noun collocates, whether in the free collocations or the restricted collocations for all the four verbs, occurred repeatedly in both learner corpora. On the other hand, the exclusive noun collocates used by one

group of learners or the other appeared to have a far greater variety with much fewer repetitions, especially in the free collocations. Cross-reference to Table 2 confirms that the learners of both groups tended to choose, in particular, the same set of restricted collocations. Since the verbs in focus are high-frequency words, the repetitious occurrences of the same noun collocates gave the impression that the two groups of learners shared a good stock of vocabulary in their written production in general.

Table 3: Occurrences of shared and exclusive noun collocates

		Free collocations			Restricted collocations		
		Exclusive CSLE	Shared	Exclusive SSLE	Exclusive CSLE	Shared	Exclusive SSLE
have	Type	20	8	25	21	11	19
	Token	32	49/38*	41	25	50/46	31
do	Type	1	1	4	4	5	9
	Token	1	11/17	4	10	11/8	15
take	Type	7	1	3	7	6	7
	Token	9	1/5	3	12	39/46	10
make	Type	2	0	2	12	4	9
	Token	9	0	2	21	19/19	12
Total	Type	30	10	34	44	26	44
	Token	51	61/60	50	68	119/119	68

*The first figure is the number of tokens in CSLE, and the second the number in SSLE.

4.4 Shared and exclusive errors

Table 4 shows the overall distribution of the wrong or problematic collocations in terms of tokens. It can be seen that in general errors of all kinds covered more than 20 per cent of each category and the figures were similar for both corpora, though on the whole there were slightly more errors made by the Chinese learners than the Swedish ones.

Table 4: Overall distribution of wrong or problematic collocations

		CSLE	SSLE
have + noun	Free	14	17
	Restricted	22	16
	All	36	33
do + noun	Free	1	3
	Restricted	9	9
	All	10	12
take + noun	Free	5	3
	Restricted	15	19
	All	20	22
make + noun	Free	9	1
	Restricted	6	6
	All	15	7
verbs + noun	Free	29	24
	Restricted	52	50
	All	81	74

Repetition of exactly the same error was rather rare. For instance, the colloca-
tions *have its responsibility to* (*have the responsibility to*), *make difference*
(*make a difference*) occurred only once in the corpora. But the type of problem
which they illustrate – problematic determiners with appropriate nouns – was
quite frequent. Therefore we classified the errors into fairly abstract types and
focused on these rather than the individual items which were problematic. All
these collocations listed above were grouped into one type and the total occur-
rence of this type was counted accordingly. There were three collocations which
involved more than one problematic type. For example, in the collocation *have
no mood in* (*be/feel in the mood for*), there are two problematic constituent parts,
namely the verb *have* and the preposition *in*. In this case, the two error types
were both considered. Thus the total occurrence of errors of different types in
Table 5 is slightly higher than that of the problematic collocations as shown in
Table 4.

Table 5: Types of errors in the verb + noun collocations

Type of error	Example	CSLE		SSLE		Total	
		FC	RC	FC	RC	FC	RC
Lexical error: verb choice	*take the problem (solve the problem)* *do a great effort (make a great effort)*	15	13	18	14	33	27
Lexical error: noun choice	*have their lines (have their strategies)* *make benefit (make a profit)*	0	0	4	0	4	0
Lexical error: adjective	*do some protecting work*	2	10	0	3	2	13
Grammatical error: noun plurality	*have more equipments (have more equipment)* *have troubles with (have trouble with)*	5	9	1	5	6	14
Grammatical error: determiner	*have limited supply of (have a limited supply of)* *have its responsibility (have the responsibility* *have the duty (have a duty)*	3	9	0	9	3	18
Grammatical error: preposition	*take heed to (take heed of)* *have difficulties to do (have difficulties in doing)* *do harm of (do harm to)*	0	6	0	9	0	15
Grammatical error: syntactic structure	*do favor to (do sb. a favor)*	0	1	0	0	0	1
Grammatical error: adverb form	*have a full functional sanitation (have a fully functional sanitation)*	0	0	1	1	1	1
Semantic error: correct collocation doesn't make sense	*have the ability to worry about it* *take care of the problems*	4	5	0	10	4	15

Table 5 displays all the error types occurring in the two corpora. Table 6 summarizes to demonstrate that most of the errors concerning the free collocations in both corpora were caused by mismatch between lexical items. Since our method

involved selecting verbs, we picked up 'wrong verbs' when the verb actually used was one of the four focused on. Therefore the errors of this type are mostly wrong choices of verb. For the restricted collocations the observed errors were mainly about the other types such as unacceptable determiners, which concern restrictions on the variability of collocations. A chi-square test reveals that the difference is highly significant. To put it in another way, the learners actually have problems with semantic categories in the use of free collocations since they typically have few restrictions on, e.g. determiners, where lie chiefly the learn-ers' difficulties with the restricted collocations. It is worth noting that both cor-pora contributed quite evenly to the distribution of the above two seemingly most problematic areas. That is to say, these major difficulties do not appear to be L1 related.

Table 6: Comparison of different types of errors

	Problematic lexical choices				Other types of errors			
	CSLE		SSLE		CSLE		SSLE	
	n.	%	n.	%	n.	%	n.	%
Free collocations	17	(32)	22	(41)	12	(23)	2	(4)
Restricted collocations	23	(22)	14	(13)	30	(29)	37	(36)

Numbers in parentheses are percentages out of the total number of the problematic error types in each category; χ^2 =28.21, df =1, p < 0.001.

5   Discussion

The essays written by the Chinese-speaking learners and the Swedish-speaking learners were different in aspects like style, grammatical mistakes, and register. For example, in terms of style, the essays written by the Swedish-speaking learners usually began directly with a topic statement, but the essays written by the Chinese-speaking learners were marked by what is called by Kaplan (1966: 10) an approach of "indirection". That is to say, instead of addressing the subject directly, the Chinese students very commonly started by alluding to something else before coming to the real subject. The following excerpt from one Chinese student's writing may serve as a typical example:

> *In a poor family, without enough money to support his wife and chil-dren, a man would never think of decorating his house or buying*

> *expensive plants to keep the air in the house fresh. If he would, his family are bound to suffer from hunger. Like a poor family, a poor country faces the same dilemma.*

This kind of writing demonstrates strong influence from the learners' Chinese thought patterns, or "the Oriental spiral", as Kaplan (1966: 10) described, which differ markedly from the linear English thought patterns. Unlike the Chinese students, almost all the Swedish students, whose L1 and culture are closely related to English, started their essays by a direct statement of their opinion towards the topic, as the following excerpt shows:

> *The claim that only rich countries can afford to worry about the environment is in fact both wrong and right to some extent.*

The difference in style of the essays written by the two groups of learners suggests the influence from the learners' L1 backgrounds, or at least such influence seems to be more obvious when the L1 and the TL are more typologically distant.

When it comes to vocabulary, according to many linguists and researchers (e.g. Blum and Levenston 1978; Ringbom 1983; Farghal and Obiedat 1995; Jiang 2004), language learners put to use as a working hypothesis that there is a word-for-word translation-equivalence between L1 and L2. Thus it was assumed that, given the subjects' quite different L1 backgrounds and different English learning experience in the present study, they might also demonstrate fairly different collocational proficiency, especially in terms of lexical choice. The present study, based on parallel learner corpora, from which was extracted a large sample of comparable collocations, allowed an empirical understanding of EFL learners' collocation use. The result of the comparison does not support the assumption. To begin with, the total occurrences of the verb + noun collocations investigated were similar in both corpora. Secondly, the lexical variety of the noun collocates for each verb was also strikingly similar in both corpora. Thirdly, both groups of students tended to make repeated use of the same set of noun collocates for each of the four verbs, especially in restricted collocations. All these seem to suggest similarity in the collocation development of learners with different L1 backgrounds. Restricted collocations, according to Nesselhauf (2003), are less often congruent than free combinations between two languages, since semantically possible combinations in learners' L1 are not always collocationally possible in the TL. Thus if collocation development were L1 dependent, one would expect more difficulty with restricted collocations. However, the fact that the similarity extends over to restricted collocations in the present study

supports the view that the process of acquisition of collocations is probably not very L1 dependent.

For some of the shared restricted collocations in which the verb and noun components were correctly combined such as *do damage*, *make a difference*, word-for-word equivalents cannot be found in either Chinese or Swedish. So the subjects' correct lexical choice in these cases is probably not due to lucky use of a word-for-word translation strategy. Errors are also similar in terms of lexical choice. Both groups confuse, in particular, *do* and *make*, although Swedish has a single verb *göra* which can be used delexically; for example, *göra förändringar* and *göra sitt bästa* in Swedish correspond to *make changes* and *do one's best* respectively in English. Some previous studies (e.g. Biskup 1992; Bahns 1993; Granger 1998b) have come to the conclusion that L1 transfer is quite likely in collocation use especially when learners' L1 and the TL are closely related. Thus it is not surprising to find in SSLE the problematic collocations like *do changes* and *make the cleaning* due to the effect of the negative L1 transfer. However, similar errors were also found in CSLE such as *do a great effort* (*make a great effort)* and *make damage* (*do damage*). The equivalents of *do* (*zuo*) and *make* (*zhizhao*) in Chinese are not used as delexicalized verbs as they often are in English. That is to say, it doesn't make any sense to combine either *do* (*zuo*) or *make* (*zhizhao*) with *effort (Jinli)* or *damage* (*pohuai*) in Chinese. Therefore in these cases it might be questionable to simply attribute such errors to L1 transfer.

The use of *make* in native and non-native student writing has been the focus for quite a few studies. Among them Altenberg and Granger (2001) compared Swedish- and French-speaking learners and Liu and Shaw (2001) involved Chinese-speaking learners. The results seem to suggest that some problems, e.g. overuse of a high-frequency verb, misuse and underuse of collocations, are universal for advanced EFL learners. The present study supports this view of universality in relation to the learners' lexical choices, both correct and erroneous, in a part of the restricted collocations.

Apart from a few frequently used free collocations across the corpora such as *do thing(s)*, *have money, have time*, a large proportion of the free collocations were used exclusively by one group of learners or the other. This is a consequence of their freedom of course. The study of errors in this relation turned out not to conform to the suggestion of Howarth (1998) that learners do well at the two end-points of prefabs, i.e. free collocations and idioms. The major problem concerning the use of free collocations is wrong lexical choice as in *make factories*, *do recycling* (SSLE) and *do the pollution* (CSLE).

On the one hand, the learners seem to be unaware of the semantic incompatibility between *make* and *factories*, probably due to the influence of their L1. On

the other hand, there is a tendency towards blending. But different from Howarth's (1998) observation that it is quite likely for learners to blend two restricted collocations, the learners in the study tended to blend free collocations such as *recycle paper/glass*, *pollute the air/water/environment* with restricted collocations such as *do the cooking/washing/shopping* and *do the laundry*. This probably accounts for the combinations such as *do recycling* and *do the pollution* occurring in the learner corpora. In a study of the lexical usage of advanced adult L2 learners of Hebrew, Levenston and Blum (1977) observed that the L2 learners tend to use more words of general than of specific meaning and to generalize inappropriately. The Chinese learners and the Swedish learners in the present study seem to have the same problem, especially given that the four delexical verbs are of the focus. Thus the problem is probably intralingual rather than interlingual.

On the whole, the Chinese learners made relatively fewer mistakes of the type of wrong choice of verb, whether in the free or the restricted collocations. Combined with the result that there was a slightly smaller variety of noun collocates in CSLE, it might be reasonable to suppose that the Chinese learners were more conservative and cautious in the choice of a noun collocate for a particular verb, whereas the Swedish learners tended to be more creative, hence generating greater variety and more mistakes likewise in this respect. Interestingly, Biskup (1992), comparing the use of collocations by Polish and German learners of English, also found that the Germans, whose L1 is closely related to English, were more inclined to take risks, whereas the Poles were more conservative, hence giving more correct answers than the Germans did. Thus there is an indirect influence of the learners' background on their IL, as far as collocations are concerned. In the case of two closely related languages, such as Swedish and English, it is possible that learners tend to over-extend the similarity between the two languages, hence being more creative in word-combination. But in the case of languages that are perceived as distant, such as Chinese and English, learners might be more cautious in the choice of a word to go with the other, which results in lack of variety, yet fewer errors in this respect. In other words, the assumptions that the learners have about the distance between their L1 and the TL might lead to different strategies concerning collocational production in the TL (Kellerman 1977). This implies that the view that the distance between the languages concerned is an important determining factor in language transfer also applies to collocation.

The correct choice of the lexical components does not necessarily mean that the learners had no problem in using the collocations correctly. Especially in the case of the restricted collocations, as shown in Table 6, other types of errors than

lexical choice cover up to two thirds of all the errors occurring in the corpora. This result confirms Nesselhauf (2003) that, to produce an acceptable verb + noun collocation, EFL learners have to know the whole combination as well as to know which lexical items collocate. The major problems, as shown in Table 5, involved correct use of determiner and number of the noun collocates. The Chinese learners made more errors of both types presumably because these categories and forms are quite different in Chinese. For instance, the Chinese students used the collocations such as *have its right to*, *have its/his/responsibility*, seemingly unaware that the determiners might be sensitive to variability in these restricted collocations. In addition, there were also cases in which the Chinese learners used plural form for some constituent parts like *part* and *damage* in the above *do* + noun collocations. Interestingly enough, they used *do its part*, but also *do our parts,* without knowing that *part* in this collocation should be an uncountable noun. Similarly, there was *do damage* in the CSLE, but also *do great damages* perhaps to express different kinds of damage done by people to the environment.

It is worth noting that, unfortunately, in the case of *do damages*, the problem casts some doubt on the notion of collocation errors. In other words, it is rather difficult to decide whether such errors are collocational or simply grammatical. It seems that the error in *do damages* is caused by simple ignorance of the grammar of *damage*; namely, we can never say *damages* with this meaning. Therefore, the problem concerning number might not really be a reflection of the learners' unawareness of the syntactic restrictions of certain collocations, but rather of general grammatical/lexical problems. There were some other examples such as *do a lot of researches* (*do a lot of research*), *have benefit* (*have benefits*), etc., and some cases concerning wrong use or absence of determiner such as *have good environment* (*have a good environment)*, all of which strongly suggest that the errors were made because the learners were confused about the distinction of countable nouns versus uncountable ones rather than the specific restrictions about the use of certain collocations. The same problem seems to exist in some previous studies. Nesselhauf (2005), for example, included errors as in *give children a sound knowledge* and *put an enormous pressure on* in the category of 'article superfluous' in collocation use without specifying the complexities of defining an error.

As for the probable reason for the misuse in the above cases, it could have something to do with the tendency for EFL learners to generalize from their previous knowledge of rules or from misconceptions of certain syntactic and semantic rules in the TL (Olsen 1999). Since there are no inflectional changes for a word in Chinese, the Chinese students in general might have more difficul-

ties in acquiring the inflectional rules of English, compared with the Swedish learners whose L1 has much in common with English, in terms of both grammar and lexicon. However, errors of this type were not confined only to the Chinese learners throughout the study. In SSLE, such mistakes as *have troubles with, have the duties, have a bigger knowledge, make difference* also occurred quite frequently. Therefore it might be true that these problems have more to do with the learners' lack of awareness of the restriction concerning the variability of the collocations, or the general syntactic rules, rather than the effect of L1 transfer. However, it is also likely that the Swedish learners, with their L1 typologically much closer to the TL, found it easier to avoid such problems. Therefore the influence of the learners' L1 in the TL collocational use is intricate in that the L1 might facilitate the learners' TL learning in some way.

To sum up, the present study confirms the claims of some researchers (e.g. Lennon 1996, Farghal and Obiedat 1995; Jiang 2004 and others) that advanced learners still have a great problem with the use of common words and collocations, yet not only restricted collocations but also free ones. Meanwhile it casts doubt on the emphasis on L1 transfer in the learners' collocational problems as maintained by some researchers (e.g. Bahns and Eldaw 1993; Nesselhauf 2003, etc.). One might usually assume that collocational problems are a type of vocabulary difficulty and hence largely L1 based, in the sense that it is easier to learn new words if their forms are related, like English and other Germanic or Romance languages, or probably if the semantic structure is similar, like English and other European languages. However, the collocation side of language learning for the learners of advanced level, as suggested in the present study, seems to be more closely related to syntax learning, which is often affected by L2-based or 'universal' features whereas the influence from the learners' L1 is found to be in a subtle fashion. As Altenberg and Granger (2001: 184) suggest, the learners' performance should be explained as the result of several factors – "interlingual, intralingual, and inadequate teaching". Therefore L1 transfer is only one of the factors. Moreover, the influence of the learners' L1 is more profound than simply gap-filling in the learners' knowledge of the TL with items from the L1. Specifically, it might facilitate the learners' acquisition of a S/FL and influence their learning and performing strategies in some way. Other factors, which are more TL related and might be universal to the development of learner's IL, such as the learner's unawareness of the restrictions concerning the syntactic variability of restricted collocations and the context in which they can be used, their misconceptions about certain syntactic and semantic rules in general, the over-extension of certain high-frequency verbs and collocational structures, etc., should also be given adequate attention and investigation.

The results have pedagogical implications concerning the instruction of high-frequency verbs, which tend to be neglected after they are encountered in an early stage, but are still repeatedly used erroneously by advanced learners. The similarity of the ranges of noun collocates in the restricted collocations in the two learner corpora shows that the learners were rather conservative in the use of restricted collocations, which suggests that the acquisition of restricted collocations might not have been totally neglected in the English instruction for both groups of learners. But considering the errors made by the learners in this area, it is also very likely that frequent collocations have been encountered incidentally from mere exposure, which is necessary (Pawley and Syder 1983), but not enough (Marton 1977), for the acquisition of collocations. On the other hand, the range of lexical choice in the free collocations was quite wide and many of the choices were wrong because of the learners' insufficient knowledge about these high frequency verbs and their delexical structures. This, again, reflects the learners' unawareness of collocational restrictions. Therefore there is an urgent need to help them through systematic instruction to appreciate this problem and raise their awareness of the complexity of collocation in relation to syntactic flexibility, appropriate contexts and so on.

The present study is only an exploratory attempt to investigate the learners' collocational problems. First of all, there are definitional problems. There is no universally accepted formal definition and classification of collocation, which might lead to problems for the methodology utilized in the study. However, the unsettled theoretical description also means that such experimental attempts are worthwhile and necessary. As we have shown, the notion of collocational error is not well defined either and results might depend on what we regard as a collocational error and what as a straightforward grammatical error or wrong lexical choice. Secondly, there are elements of subjectivity and arbitrariness in our classifications. Despite strenuous work and long hours spent deciding the acceptability of the collocations, some decisions are open to discussion. Similarly, the interference of the learners' L1 on their use of collocations in the present study was judged relying solely on the authors' intuition. In an ideal world in order to decide whether the occurrence of a specific collocation is due to the interference of the learners' L1, we would work from corpora of the two sets of learners' L1 and a TL corpus consisting of native English speakers' essays written about the same topic so that deviations of the learners' ILs from the target norm could be uncovered. Through a detailed cross comparison of the learners' IL, their different L1 and the TL by the use of contrastive corpus analysis, a better understanding of the interference of L1 transfer and the influence of the other factors as mentioned in the present study on the development of learners' IL might be achieved.

6   Conclusion

It has been widely accepted that "words are very good candidate for units of equivalence" (Selinker 1992: 32) across linguistic systems so we assumed that we could easily detect the effect of EFL learners' L1 on their use of collocations. However, the development of IL collocational proficiency among the two sets of advanced English learners as reflected in the present study turned out to be quite similar in many respects despite the very different relationship between their respective L1 and the TL. Whereas the interference of the learners' L1 plays an undeniable role in the use of collocations, it is suggested that the influence is exerted in a profound way. That is to say, the correct or problematic use of the collocations investigated is not simply caused by the learners' searching for equivalent "candidates" across the linguistic systems. But the different relationship between the learners' L1 and the TL might lead to the different strategies adopted by them in collocational production, and the learners' TL learning might be facilitated in some way if their L1 is closely related to the TL. In addition, there are some other important factors that are TL based and might be universal to all the learners in the development of IL collocational proficiency but have yet not been given adequate attention in SLA theory. The fact that ILs are shown to be at least partly systematic means that "they are potentially amenable to systematic change, e.g. through instruction" (Larsen-Freeman and Long 1991: 83). Therefore, in further studies it would be of great value to investigate in greater depth the universal factors which might have a strong impact on the systematic development of collocations.

Note

1. In this paper we try to avoid assuming that a highly proficient user is in fact a native speaker, recognizing that speakers who have another mother tongue may have had the exposure necessary to develop idiomatic proficiency. Most of the literature, does however, naturally enough, identify perfect knowledge and nativeness.

References

Aitchison, Jean. 2003. *Words in the mind: An introduction to the mental lexicon* (3rd ed.). Oxford: Blackwell.

Allerton, David J. 1984. Three (or four) levels of word cooccurrence restriction. *Lingua* 63 (1): 17–40.

Altenberg, Bengt. 1993. Recurrent verb-complement constructions in the London-Lund Corpus. In J. Aarts, P. Haan and N. Oostdijk (eds.). *English language corpora: Design, analysis and exploitation. Papers from the Thirteenth International Conference on English Language Research on Computerized Corpora, Nijmegen 1992*, 227–245. Amsterdam: Rodopi.

Altenberg, Bengt and Sylviane Granger. 2001. The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics* 22 (2): 173–195.

Bahns, Jens. 1993. Lexical collocations: A contrastive view. *ELT Journal* 47 (1): 56–63.

Bahns, Jens and Moira Eldaw. 1993. Should we teach EFL students collocations? *System* 21: 101–114.

BBI = *The BBI dictionary of English word combinations*. 1997. Ed. by Morton Benson, Evelyn Benson and Robert Ilson. Amsterdam and Philadelphia: John Benjamins.

Beheydt, Ludo. 1987. The sementisation of vocabulary in foreign language learning. *System* 15 (1): 55–67.

Biskup, Danuta. 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*, 85–93. Basingstoke: Macmillan.

Blum, Shoshana and Edward A. Levenston. 1978. Universals of lexical simplification. *Language Learning* 28 (2): 399–415.

Bogaards, Paul. 2000. Testing L2 vocabulary knowledge at a high level: The case of the Euralex French tests. *Applied Linguistics* 21 (4): 490–516.

Bogaards, Paul. 2001. Lexical units and the learning of foreign language vocabulary. *Studies in Second Language Acquisition* 23 (3): 321–343.

Bolinger, Dwight L. 1976. Meaning and memory. *Forum Linguisticum* 1 (1): 1–14.

Cantos, Pascual and Aquilino Sánchez. 2001. Lexical constellations: What collocates fail to tell. *International Journal of Corpus Linguistics* 6 (2): 199–228.

Carter, Ronald and Michael McCarthy. 1988. *Vocabulary and language teaching*. London and New York: Longman.

CCED = *Collins Cobuild English dictionary*. 1995. London: Harper Collins.

Chan, Alice Y. W. 2004. Syntactic transfer: Evidence from the interlanguage of Hong Kong Chinese ESL learners. *The Modern Language Journal* 88 (1): 56–74.

Cobb, Tom. 2003. Analyzing late interlanguage with learner corpora: Québec replications of three European studies.

Collier, Alex. 1993. Issues of large-scale collocational analysis. In J. Aarts, P. Haan and N. Oostdijk (eds.). *English language corpora: Design, analysis and exploitation*, 289–298. Amsterdam: Rodopi.

Cowie, Anthony P. 1981. The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics* 2 (3): 223–235.

Erman, Britt and Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text* 20 (1): 29–62.

Farghal, Mohammed and Hussein Obiedat. 1995. Collocations: A neglected variable in EFL. *IRAL – International Review of Applied Linguistics in Language Teaching* 33 (4): 315–332.

Firth, John Rupert. 1957. *Papers in linguistics.* London: Oxford University Press.

Granger, Sylviane. 1998a. The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (ed.). *Learner English on computer*, 3–18. London and New York: Longman.

Granger, Sylviane. 1998b. Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (ed.). *Phraseology: Theory, analysis, and applications*, 145–160. Oxford: Oxford University Press.

Grant, Lynn and Laurie Bauer. 2004. Criteria for re-defining idioms: Are we barking up the wrong tree? *Applied Linguistics* 25 (1): 38–61.

Henriksen, Birgit. 1999. Three dimensions of vocabulary development. *Studies in Second Language Acquisition* 21 (2): 303–317.

Hill, Jimmie. 2001. Revising priorities: From grammatical failure to collocational success. In Michael Lewis (ed.). *Teaching collocation: Further developments in the lexical approach*, 47–69. London: Language Teaching Publications.

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

Javis, Scott, Leslie Grant, Dawn Bikowski and Dana Ferris. 2003. Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing* 12 (4): 377–403.

Jiang, Nan. 2004. Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal* 88 (3): 416–432.

Judd, Elliot L. 1978. Vocabulary teaching and TESOL: A need for reevaluation of existing assumptions. *TESOL Quarterly* 12 (1): 71–76.

Källkvist, Marie. 1999. *Form-class and task-type effects in learner English: A study of advanced Swedish learners.* Lund: Lund University Press.

Kaplan, Robert B. 1966. Cultural thought patterns in inter-cultural education. *Language Learning* 16: 1–20.

Kellerman, Eric. 1977. Towards a characterization of the strategy of transfer in second language learning. *Interlanguage Studies Bulletin* 2: 58–145.

Larsen-Freeman, Diane and Michael H. Long. 1991. *An introduction to second language acquisition research*. London and New York: Longman.

Laufer, Batia. 1990. Why are some words more difficult than others? – some intralexical factors that affect the learning of words. *IRAL – International Review of Applied Linguistics in Language Teaching* 28 (4): 293–307.

Laufer, Batia. 1998. The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics* 19 (2): 255–271.

Laufer, Batia and Paul Nation. 1995. Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics* 16 (3): 307–322.

Lennon, Paul. 1996. Getting 'easy' verbs wrong at the advanced level. *IRAL – International Review of Applied Linguistics in Language Teaching* 34 (1): 23–36.

Levenston, Edward A. and Shoshana Blum. 1977. Aspects of lexical simplification in the speech and writing of advanced adult learners. In E. Roulet (ed.). *Actes du Cinquième Colloque de Linguistique Appliquée de Neuchâtel*, 51–71. Geneva: Droz.

Lewis, Michael. 2001. Language in the lexical approaches. In Michael Lewis (ed.). *Teaching collocation: Further developments in the lexical approach*, 126–154. London: Language Teaching Publications.

Lewis, Morgan. 2001. There is nothing as practical as a good theory. In Michael Lewis (ed.). *Teaching collocation: Further developments in the lexical approach*, 10–27. London: Language Teaching Publications.

Liszka, Sarah A. 2004. Exploring the effects of first language influence on second language pragmatic processes from a syntactic deficit perspective. *Second Language Research* 20 (3): 212–231.

Liu, Eric T. K. and Philip M. Shaw. 2001. Investigating learner vocabulary: A possible approach to looking at EFL/ESL learners' qualitative knowledge of the word. *IRAL – International Review of Applied Linguistics in Language Teaching* 39 (3): 171–194.

Marton, Waldemar. 1977. Foreign vocabulary learning as problem No. 1 of language teaching at the advanced level. *Interlanguage Studies Bulletin* 2: 33–57.

Nation, I. S. Paul. 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nesselhauf, Nadja. 2003. The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics* 24 (2): 223–242.

Nesselhauf, Nadja. 2005. *Collocations in a learner corpus*. Amsterdam: John Benjamins.

Odlin, Terence. 1989. *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Olsen, S. 1999. Errors and compensatory strategies: A study of grammar and vocabulary in texts written by Norwegian learners of English. *System* 27 (2): 191–205.

Partington, Alan. 1998. *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam and Philadelphia: John Benjamins.

Pawley, Andrew and Frances Hodgetts Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards and R. W. Schmidt (eds.). *Language and communication*, 191–225. London: Longman.

Ringbom, Håkan. 1983. Borrowing and lexical transfer. *Applied Linguistics* 4: 201–212.

Ringbom, Håkan. 1998. Vocabulary frequencies in advanced learner English: A corpus-linguistic approach. In S. Granger (ed.). *Learner English on computer*, 41–52. London and New York: Longman.

Reid, Joy. 1990. Responding to different topic types: A quantitative analysis from a contrastive rhetoric perspective. In B. Kroll (ed.). *Second language writing*, 191–210. Cambridge: Cambridge University Press.

Scott, Mike. 1996. *WordSmith Tools*. Oxford: OUP. Available at http://www.oup.com/elt/global/isbn/6890/

Selinker, Larry. 1992. *Rediscovering interlanguage*. London and New York: Longman.

Sinclair, John. 1991. *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Singleton, David. 2000. *Language and the lexicon: An introduction*. London: Arnold.

Stubbs, Michael. 2001. *Words and phrases: Corpus studies of lexical semantics*. Oxford and Massachusetts: Blackwell.

Wallace, Michael J. 1984. *Teaching vocabulary.* London: Heinemann Educational Books.

Wang, Lurong. 2003. Switching to first language among writers with differing second-language proficiency. *Journal of Second Language Writing* 12 (4): 347–375.

Appendix I: The first 20 verbs in the raw wordlists of each learner corpus

	Wordlist CSLE			Wordlist SSLE		
	Word-form	Frequency	%	Word-form	Frequency	%
1	IS	484	2.01	IS	358	1.79
2	CAN	334	1.39	HAVE	270	1.35
3	HAVE	210	0.87	CAN	223	1.12
4	ARE	207	0.86	ARE	166	0.83
5	WORRY	199	0.83	DO	154	0.77
6	AFFORD	190	0.79	BE	143	0.72
7	SHOULD	157	0.65	WORRY	107	0.54
8	DO	125	0.52	AFFORD	94	0.47
9	BE	110	0.46	THINK	92	0.46
10	PROTECT	106	0.44	WILL	85	0.43
11	THINK	105	0.44	DON'T	73	0.37
12	WILL	92	0.38	IT'S	73	0.37
13	IT'S	68	0.28	CARE	67	0.34
14	HAS	66	0.27	WOULD	65	0.33
15	TAKE	60	0.25	HELP	58	0.29
16	DON'T	53	0.22	SHOULD	58	0.29
17	MAKE	47	0.20	TAKE	58	0.29
18	PAY	47	0.20	MAKE	51	0.26
19	CAN'T	46	0.19	HAS	44	0.22
20	DEVELOP	34	0.14	COULD	40	0.20

Appendix II: Common and exclusive noun collocates of the verbs under study and their occurrences in the learner corpora (exclusive nouns with frequency less than 2 omitted)

(a) Common and exclusive noun collocates of have *in the learner corpora*

Exclusive CSLE	Common	Exclusive SSLE
earth 10	money 27/27	possibility(ies) 8
capacity(ies) 3	responsibility(ies) 24/14	opportunity(ies) 3
support 2	ability 8/1	knowledge 3
doubt 2	time 6/2	education 3
factory(ies) 2	right(s) 6/2	funding(s) 3
equipment 2	problem(s) 3/7	economy 2
idea (of) 2	advantage(s) 3/3	world 2
influence 2	industry(ies) 3/6	technology 2
	food 1/1	weapons 2
	car(s) 1/5	harvest 2
	environment 3/1	obligation 2
	thing(s) 1/4	
	effect 3/2	
	difficulties 2/1	
	power 2/2	
	capability 1/1	
	rule 1/1	
	choice 1/3	
	duty 2/2	

(b) Common and exclusive noun collocates of do *in the learner corpora*

Exclusive CSLE	Common	Exclusive SSLE
harm 6	thing(s) 11/17	(one's) bit 4
effort(s) 2	(one's) best 4/1	mistake(s) 3
favor 1	damage 1/4	changes 1
job	(one's) part 2/2	choices
pollution	research 2/1	consumption
	(one's) work 2/1	experiments
		improvments
		journey
		plans
		recycling
		tests
		the same
		travelling

(c) Common and exclusive noun collocates of take *in the learner corpora*

Exclusive CSLE	Common	Exclusive SSLE
(China) for example 3	responsibility(ies) 13/11	car 2
status 2	care 2/25	place 2
charge (of) 2	measure(s) 14/2	a look (at) 2
way(s) 2	problem 1/5	aim 1
a turn 2	step(s) 3/3	bus
(the environment) into consideration 2	action(s) 6/1	concept
advantage 1	part 1/3	heed
bag		jars
batteries		research
entertainment		transportation

(economy) in the first place

(the environment) as a sacrifice

methods

rubbish

(d) Common and exclusive noun collocates of make *in the learner corpora*

Exclusive CSLE	Common CSLE/SSLE	Exclusive SSLE
pollution(s) 8	effort(s) 15/8	mistake(s) 4
use (of) 7	difference 2/8	benefit 1
progress 3	change(s) 1/2	cleaning
laws 2	money 1/1	contribution
achievements 1		decisions
analysis		factories
damage		omelette
holes		restrictions
inspection		sacrifices
plan		the best (out of)
point		war
policies		weapons
problem		
project		