PREFABRICATED PATTERNS IN ADVANCED EFL WRITING: COLLOCATIONS AND LEXICAL PHRASES

Sylviane Granger

DRAFT

in: A.P. Cowie (ed.) (1998) *Phraseology: Theory, Analysis and Applications*. Clarendon Press: Oxford, 145-160.

0. Introduction

Over the last decade, the use of prefabs has become a major focus of interest in EFL, arguably for three main reasons. Firstly, the emergence of the concept of lexico-grammar, inspired by Halliday and Sinclair, has promoted the syntagmatic investigation of lexis. The traditional association between syntagmatics and grammar, on the one hand, and paradigmatics and lexis, on the other, is a thing of the past.

Secondly, corpus linguistics has played an important role, giving linguists the computational means to uncover and analyse lexical patterns. Rich information about word combinations can now be obtained with ease using text retrieval software.

Finally, pragmatics has become a major field of study in its own right, in linguistics and now in EFL. Pragmatic competence has come to be viewed as an essential part of learners' competence. The formulaic nature of many pragmalinguistic rules has necessarily contributed to bringing the study of prefabs to the fore.

1. Prefabs in learner writing

Work at Louvain on word combinations was inspired by the **International Corpus of Learner English** (ICLE) project, a project whose aim was to gather and computerize a corpus of EFL writing from learners of various mother tongue backgrounds (see S. GRANGER (1993). Although the corpus is not yet complete, research already undertaken in the fields of lexis and discourse has demonstrated its potential to uncover factors of non-nativeness in advanced learners' writing.

The methodology employed for most of this research I have termed Contrastive Interlanguage Analysis (CIA), (see S. GRANGER forthcoming). CIA may involve two types of comparison: a comparison of native and non-native varieties of one and the same language: L1 vs L2, or a comparison of several non-native varieties: L2 vs L2.

The investigations presented in this paper are based on the former type of comparison. The initial hypothesis was that learners would make less use of prefabs, or conventionalised language, in their writing than their native speaker counterparts, given that the use of such language is universally presented as typically native-like. I hypothesised that learners would make much greater use of what J. SINCLAIR (1987, 319) calls the **open choice principle** than native speakers, who have been found to operate primarily according to the **idiom principle**. To use G. KJELLMER's (1991, 124) metaphor, I expected the learners' building material to be individual bricks rather

than prefabricated sections. The data compared came from a corpus of native English writing and a similar corpus of writing by advanced French-speaking learners of English. The learner corpus is a subcorpus of the ICLE database. The native speaker corpus is made up of three main parts: the Louvain essay corpus, the student essay component of the International Corpus of English (ICE) and the Belles Lettres category of the LOB corpus[1].

2. Two types of prefabs: collocations and lexical phrases

For the purposes of the investigation, a distinction is made between collocations and lexical phrases. The term collocation is used to refer to "the linguistic phenomenon whereby a given vocabulary item prefers the company of another item rather than its 'synonyms' because of constraints which are not on the level of syntax or conceptual meaning but on that of usage" (J. VAN ROEY, 1990, 46). This phenomenon is illustrated in combinations such as **commit suicide, sound asleep** or **pitched battle,** which M. BENSON et al (1986) call 'lexical collocations' and E. AISENSTADT (1979) 'restricted collocations'. Following NATTINGER & DECARRICO (1992) I oppose this type of string to lexical phrases, such as **be that as it may** or **it seems (to me) (that) X**, which have pragmatic functions.

2.1 Collocations

2.1.1. Collocational study of amplifiers

For the collocational study, one category of intensifying adverbs was selected: amplifiers ending in **-ly** and functioning as modifiers, such as those in examples 1-3 below:

(1) ...although this feeling is perfectly natural.
(2) ...themes in Les Mouches which are very closely linked with
(3) ...a young man who is deeply in love.

They constitute a particularly rich category of collocation, involving as they do a complex interplay of semantic, lexical and stylistic restrictions and covering the whole collocational spectrum, ranging from restricted collocability - as in **bitterly cold** - to wide collocability - as in **completely different/new/free/etc.** In including adverbs such as **bitterly** in **bitterly cold** or **unbearably** in **unbearably ugly**, I have adopted a much wider notion of amplifier than other linguists such as U. BACKLUND (1973), who rejects adverbs such as these which express both degree and manner.

Using the text retrieval software TACT, all the words ending in **-ly** were automatically retrieved from the native and non-native corpora and then manually sorted according to the predefined semantic and syntactic criteria.

As a first step, the number of tokens and types in the two corpora were compared, revealing a statistically very significant underuse of amplifiers in the learner corpus, both in the number of tokens and types (see Table 1).

Table 1: Raw frequencies of amplifiers based on a NS corpus of 234,514 words and a NNS corpus of 251,318 words

	NS	NNS
Types	**75**	**41** $^{-**}$
Tokens	**313**	**230** $^{-**}$

The next step was to establish whether this underuse was general or due to underuse of particular amplifiers or categories of amplifiers.

Of the individual amplifiers, only three demonstrated statistically significant differences, as shown in Table 2: **completely** and **totally** were overused by the learners and **highly** was underused. On the whole, however, the frequencies of the individual amplifiers were often too low to draw meaningful conclusions.

Table 2: Raw frequencies of <u>completely</u>, <u>totally</u> and <u>highly</u> in the NS and NNS corpus

	NS	NNS
completely	**15**	**42** $^{+**}$
totally	**18**	**46** $^{+**}$
highly	**31**	**11** $^{-**}$

The wide variety of words with which the learners combined **completely** and **totally** - 36 different collocates for **completely**, 34 for **totally** - suggested that they are used as 'all-round amplifiers' or safe bets. Indeed, practically none of the combinations produced were felt to be unacceptable or even awkward by native speakers. One possible explanation for their overuse may well be that they have direct translational equivalents which are very frequent in French, **complètement** and **totalement**, and which display similarly few collocational restrictions.

There may be an equally feasible interlingual explanation for learners' underuse of **highly**, whose literal equivalent, **hautement**, is only used in formal language and is relatively much less frequent. It is striking to note that the few combinations that the learners actually used - such as **highly developed** / **civilized** / **specialized** / **probable** - translate very nicely into French.

When it came to examining the amplifiers by category, I chose to apply R. QUIRK et al's (1985, 590) categorisation of amplifiers into maximizers and boosters. Maximizers are amplifiers such as **absolutely, entirely, totally,** which express the

highest degree, while boosters, such as **deeply, strongly, highly,** merely express a high degree.

Table 3: Raw frequencies of maximizers and boosters in the NS and NNS corpus

	Types		Tokens	
	NS	**NNS**	**NS**	**NNS**
Maximizers	**10**	**10**	**106**	**150**
Boosters	**65**	**31** $^{-**}$	**207**	**80** $^{-**}$
Total	**75**	**41** $^{-**}$	**313**	**230** $^{-**}$

As shown in Table 3, learners used the same number of types and a slightly higher number of tokens (mainly due to overuse of **completely** and **totally**) in the maximiser category, but the overall figures are not statistically significant. However, the categorisation revealed an underuse of boosters by the learners significant enough to explain the general underuse of amplifiers attested to earlier.

The category of boosters represents 66% of the amplifiers in the NS corpus vs only 35% in the learner corpus and the number of types is much higher than in the category of maximizers, understandably, given that boosters represent an open-ended set. Quoting examples such as **admirably fair** or **dazzlingly clear**, BOLINGER (1972,25) has pointed out that "virtually any adverb modifying an adjective tends to have or to develop an intensifying meaning".

Further subdividing the boosters into three categories -those that are exclusively used by the natives, those that are exclusively used by the learners and those that occur in the two corpora (see Table 4) - revealed further insights into the difference in the use of boosters between natives and learners.

Table 4: Boosters: types exclusive to natives or learners and types common to both.

	NS-only	NS/NNS	NNS-only
Learner corpus		**24 (77.5%)**	**7 (22.5%)**
Native corpus	**41 (63%)**	**24 (37%)**	

The majority of the NNS boosters (77.5%) were used by native speakers too, while the majority of the NS boosters (63%) were used exclusively by natives.

Broadly speaking, the native-exclusive combinations[2] fell into two categories: stereotyped combinations such as **acutely aware, keenly felt, painfully clear, readily available, vitally important** and creative combinations such as **ludicrously ineffective, monotonously uneventful, ruthlessly callous, astonishingly short**.

Both types of combination were significantly underused by the learners. The learner corpus contained some rare examples of creative combinations, such as **ferociously menacing**, **shamelessly exploited**, but these were not always very successful: **dangerously threatened** and **irretrievably different** might seem odd to a native speaker.

Interestingly, the few stereotyped combinations used by the learners typically have a direct translational equivalent in French, or are 'lexically congruent' to use J. BAHNS's (1993) terminology. For example, **closely** and its French equivalent **étroitement** have very similar collocational ranges, as do **deeply** and **profondément**. Several of the combinations used by the learners are typical combinations both in English and in French: **closely linked, closely related, deeply moved, deeply convinced, deeply rooted, deeply hurt** (see Table 5). The collocation **deeply rooted**, for instance, which occurs 8 times in the learner corpus, corresponds to **profondément enraciné**, which is mentioned as a typical combination in most French dictionaries.

Table 5: NS and NNS collocations with closely, deeply and severely

	NS	**NNS**
closely	linked (4) integrated attached	linked (3) involved related
deeply	moved convinced affected	moved convinced rooted (8) hurt in loved changed divided
severely	punished restricted shaken attacked depleted complicated felt flogged	punished

For **severely,** the case is particularly striking: of all the combinations used by the natives, the only one that translates into French is precisely that used by the

learners: **severely punished** which corresponds to **sévèrement puni**. All the other combinations used by the natives would be impossible in French: **sévèrement restreint/ébranlé/attaqué/diminué/etc.**

There is also evidence that learners use non-congruent combinations, albeit comparatively few of them. In fact, there are only three obvious ones, and these are **badly injured**, **finely detailed**, and **widely held**.

At this stage then, the investigation supports the initial hypothesis that learners use fewer prefabs than their native speaker counterparts. Further, there is evidence that the collocations used by the learners are for the most part congruent and may thus result from transfer from L1.

But the general picture is one of learners who seem to use amplifiers more as building bricks than as prefabricated sections. They tend to use some amplifiers as 'all-rounders', a tendency confirmed by their use of the amplifier **very** which although not part of the present investigation was analysed independently. The analysis showed a highly significant overuse of **very**, the all-round amplifier par excellence[3]. From the figures in Table 6 one could postulate that the learners' underuse of **-ly** amplifiers is compensated for by their overuse of **very**.

Table 6: Relative frequencies of -ly amplifiers and very based on 200,000 words per variety

	NS	NNS
-ly amplifiers	**267**	**183** $^{-**}$
very	**190**	**329** $^{+**}$

2.1.2. Significant collocation

It has been established above that learners are using collocations, but that they underuse native-like collocations and use atypical word combinations. The results of an independent study we carried out suggest that this is probably due to an underdeveloped sense of salience and of what constitutes a significant collocation.

The aim of this study was to extract introspective data on collocations and involved submitting a word combination test to 112 informants, 56 French learners and 56 native speakers[4]. Informants were asked to choose the acceptable collocates of 11 amplifiers, from a list of 15 adjectives, by circling all the adjectives which in their opinion collocated with the amplifier. If they were unsure about a particular adjective, they were instructed to underline it and if they felt that one adjective was more frequently associated with the amplifier than all the others, they were requested to mark it with an asterisk.

It was the comparison of the forms marked with an asterisk by the learners and

the natives, and which therefore indicated those combinations which were particularly salient in the subjects' minds, that yielded particularly interesting results.

All in all the learners marked with an asterisk over 100 fewer combinations than the natives (280 vs 384). Table 7 gives clear evidence of the learners' weak sense of salience. **Readily available**, for instance, was asterisked by 43 native speakers but by a mere 8 learners. **Bitterly cold** was selected by 40 native speakers and only 7 learners. For **blissfully,** the native speaker selections were evenly distributed between **blissfully happy** and **blissfully ignorant,** asterisked by 19 and 20 informants respectively, while not one single learner marked the latter combination and only 4 selected the former.

Table 7: Native speaker and learner responses to word-combining test

NS	NNS
readily available (43)	readily available (8)
bitterly cold (40)	bitterly cold (7) bitterly aware (3) bitterly miserable (2)
blissfully happy (19) blissfully ignorant (20)	blissfully happy (4)
fully aware (33) fully reliable (3)	fully aware (21) fully reliable (15) fully different (6) fully significant (5) fully impossible (3) fully available (2)
highly significant (33) highly reliable (3) highly important (2) highly aware (3)	highly significant (15) highly reliable (7) highly important (6) highly impossible (6) highly difficult (5) highly essential (4) highly different (2)

On balance, the learners marked a greater number of types of combinations than the natives, indicating that the learners' sense of salience is not only weak, but also partly misguided. Although there was evidence of a good sense of salience among a significant number of learners for some combinations, such as **fully aware**, and **fully reliable**, the learners also considered 4 other combinations to be significant collocations, none of which were selected by the native speakers: **fully different** / **significant** / **impossible** / **available.** Besides selecting **highly significant,** learners also marked six other combinations with **highly**, four of which were not marked by

native speakers. In fact, **highly impossible** / **difficult** / **essential** / **different** were together selected more often than **highly significant**. This is somewhat paradoxical when considered in light of the fact that learners underuse **highly** in their writing, but this could perhaps be put down to the production/reception distinction.

Aside from demonstrating that introspective data can play a role in revealing features of learner language, the study also suggests that this type of test could be valuable in providing a clearer notion of what constitutes a significant collocation. Certainly, there is a problem with using corpus data. As J. CLEAR (1993, 274) says: "By far the majority of lexical items have a relative frequency in current English of less than 20 per million. The chance probability of such items occurring adjacent to each other diminishes to less than 1 in 2,500,000,000! Reliable evidence of patterning between such items can be obtained *only from very substantial text corpora* .." (my italics). This is supported by the fact that G. KJELLMER's (1994) dictionary of collocations, based on the one million word Brown corpus, does not contain some common combinations, such as **blissfully happy, highly significant** or **seriously ill.**

2.2 Lexical phrases: sentence builders

J. NATTINGER & J. DECARRICO (1992, 1) define lexical phrases as "multi-word lexical phenomena that exist somewhere between the traditional poles of lexicon and syntax, conventionalized form/function composites that occur more frequently and have more idiomatically determined meaning than language that is put together each time". Such phrases are, in their opinion, a pervasive phenomenon in both speech and writing. As research at Louvain focuses on learner writing, I chose to investigate lexical phrases in writing, examining in particular the category of sentence builders, phrases which function as macro-organizers in the text, a study which fits in well with wider research being conducted at Louvain into coherence in learners' writing.

The study is based on two discourse frames - one passive, the other active - which are used to state the discourse purpose. Both frames are instances of what A. PAWLEY (this volume) calls productive speech formulas, i.e. constructions whose lexical content is only partly specified. A precise description of these two frames is given below:

Passive frame

it + (modal) + passive verb (of saying/thinking) + **that**-clause
Examples: **it is said/thought that..; it can be claimed/assumed that...**

Active frame

I or **we/one/you** (generalized pronoun) + (modal) + active verb (of saying/thinking) + **that**-clause.
Examples: **I maintain/claim that...; we can see/one could say that...**

Every instance of the pronouns **it/ I/we/you/one** followed by **that** within a span of 1 to 5 words was taken from the two corpora used in the collocational study and the relevant active and passive structures selected. The results are presented in Table 8.

Table 8: Relative frequencies of the passive and the active frame based on 200,000 words per variety

	NS	**NNS**
Passive frame	**77**	**52**
Active frame		
we/you/one	56	269 $^{+**}$
I	53	130 $^{+**}$
Total	**109**	**399** $^{+**}$

The results were most striking. While the learners made a similar use of the passive structure - both quantitatively and qualitatively -, they massively overused the active structure (c. 400 vs c. 100). Some of the frequently recurring chunks in the learner corpus are listed below. Two of the most striking examples of overuse were chunks with **say** - used 75 times by the learners but only 4 times by the native speakers and chunks with **think** - 72 in the learner corpus compared with only 3 in the native speaker corpus[5]. **Notice** and **not forget** were not used at all by the native speakers. Here again the reason for the overuse may be partly interlingual. French uses many more phatic introductory phrases than English. Phrases such as **we can say that** fill exactly the same function as **actually** or **as a matter of fact**, which have also been found to be overused by French learners in a study of connector usage in native and nonnative writing (cf. S. GRANGER & S. TYSON forthcoming).

Active frame: some recurring phrases in the learner corpus

- **we/one/you can/cannot/may/could/might <u>say</u> that:** 75 occurrences (vs <u>4</u> in NS corpus)
- **I <u>think</u> that:** 72 occurrences (vs <u>3</u> in NS corpus)
- **we/one can/could/should/may/must <u>notice</u> that:** 16 occurrences (vs <u>no</u> occurrences in NS corpus)
- **we/one may/should/must <u>not forget</u> that:** 13 occurrences (vs <u>no</u> occurrences in NS corpus)

Clearly then, while the foreign-soundingness of learners' productions has

generally been related to the lack of prefabs, it can also be due to an excessive use of them. Examples (1) to (3) below give evidence of the kind of verbosity this causes. In the three examples, two other lexical phrases, **the fact that**[6] and **as far as X is concerned**, are also underlined: these have also been found to be overused by learners and increase the impression of verbosity.

(1) Opinions are divided on this question, but as far as I am concerned **I truly believe that** this task can only be performed by each student individually.

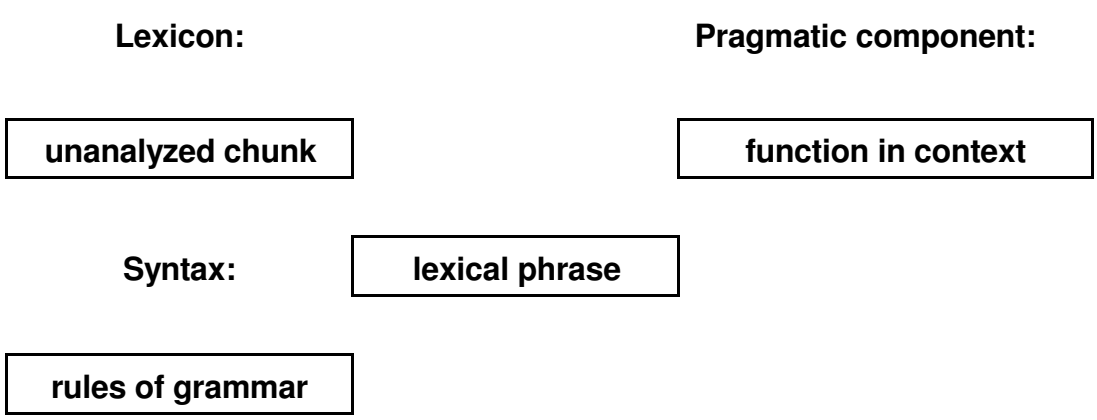
(2) I said unfortunately because **I think that** the fact that TV has too much importance for some has many bad consequences.

(3) As a conclusion, **I would say that** we cannot deny the fact that university degrees are more theoretical than practical but **I think that** it is too easy to deduce that degrees are of little value.

The use of all these phrases and frames could be viewed in terms of what H. DECHERT (1984, 227) calls "islands of reliability" or "fixed anchorage points i.e. prefabricated formulaic stretches of verbal behaviour whose linguistic and paralinguistic form and function need not be 'worked upon'". In other words, learners' repertoires for introducing arguments and points of view are very restricted and they therefore "cling on" to certain fixed phrases and expressions which they feel confident using.

3. Pedagogical implications

Conscious of the importance of prefabricated patterns in language, several EFL specialists have advocated a teaching method based on the pattern of L1 acquisition, which J. NATTINGER & J. DECARRICO (1992, 12) represent by the diagram below:



In L1 acquisition the child first acquires chunks and then progressively analyses the underlying patterns and generalizes them into regular syntactic rules. D. WILLIS (1990, iii) suggests following the same pattern in SLA, i.e. exposing learners to the commonest patterns and then relying on "the innate ability of learners to recreate for

themselves the grammar on the basis of the language to which they are exposed".

A word of caution is necessary here. It is undoubtedly important to lay greater emphasis on prefabs in ELT, especially in the case of EFL learners who have very little exposure to L1, but it seems dangerous to overemphasize the role of prefabs in SLA as research in this field is very much in its infancy. S. KRASHEN & R. SCARCELLA (1978) have surveyed several investigations into the part played by routine patterns in the development of syntactic structures both in first and second language acquisition and it is quite clear from this survey that the results are very inconclusive. If anything, the studies seem to indicate that the two strategies - routines and creative constructions - develop independently of each other and this view is supported by neurolinguistic evidence: automatic speech has been proved to be neurologically different from creative language. Within the context of L1 acquisition, A. PETERS (1977) demonstrates that children use two learning strategies: 'analytic' ("from the parts to the whole") and 'gestalt' ("from the whole to the parts") and suggests that domination of one strategy or the other will depend on individual personality and context of use.

There is very little data for adult L2 acquisition. The only investigation reported by S. KRASHEN & R. SCARCELLA (1978, 295), namely HANANIA & GRADMAN (1977), shows that the routines used by adult L2 learners resist segmentation. In other words, gestalt language fails to develop into analytic language. A more recent study by C. YORIO (1989, 68) points in the same direction: " Unlike children, they {adult L2 learners} do not appear to make extensive early use of prefabricated, formulaic language, and when they do, they do not appear to be able to use it to further their grammatical development". In other words, there does not seem to be a direct line from prefabs to creative language or to use J. SINCLAIR's (1987) terms from the idiom principle to the open choice principle. It would thus be a dangerous gamble to believe that it is enough to expose L2 learners to prefabs and grammar will take care of itself[7]. While research into the role of prefabs in L2 acquisition remains inconclusive, it seems wise to advise course designers not to overstress phraseological skills at the expense of creative skills.

Nevertheless, prefabs certainly need to play a greater role in EFL than they have in the past. The investigations presented in this paper demonstrate that learners' phraseological skills are severely limited: they use too few native-like prefabs and too many foreign-sounding ones. But if we are to devise the "ideal" pedagogical tools, a great deal more empirical data on prefabs is required. J. RICHARDS (1983, 115) considers that "many of the conventionalized aspects of language are amenable to teaching" but he adds that "applied linguistic effort is needed to gather fuller data on such forms (through discourse analysis and frequency counts, for example) with a view to obtaining useful information for teachers, textbook writers, and syllabus designers".

I suggest we need the following three types of data:

1) Detailed descriptions of English prefabricated language. The existence of computer corpora makes the compilation of collocation dictionaries possible. G.

KJELLMER's (1994) 3-volume work is the first major dictionary of this kind and makes a valuable contribution to the description of English prefabs. However, more work of this type using the new gigantic corpora is essential, if we wish to draw up lists of statistically significant collocations. As for lexical phrases, J. NATTINGER & J. DECARRICO (1992, 174) stress the need for additional empirical fieldwork and M. LEWIS (1993, 132) is equally adamant that "A resource book of lexical phrases, including sentence heads and institutionalised utterances, should be an important priority for one of the major publishing houses".

However, this type of data alone does not suffice. Learners clearly cannot be regarded as 'phraseologically virgin territory': they have a whole stock of prefabs in their mother tongue which will inevitably play a role - both positive and negative - in the acquisition of prefabs in L2. The influence of L1 routines has been brought out by psycholinguistically-oriented investigations of L2 speech production. In his description of learners' communication strategies, M. RAUPACH (1983, 208) notes that "many factors that constitute a learner's fluency in his L1 are liable to occur, in one form or another, in the learner's L2 performance" while D. MOHLE & M. RAUPACH (1989, 213) stress the complexity of L2 processing "where the learner's L2 procedural knowledge is activated in combination with parts of transferred L1 procedural knowledge".

It is thus necessary to have access to two other types of data: contrastive data and learner data, which will allow us to select the most useful prefabs for teaching purposes.

2) Good descriptions of prefabricated language in the learners' mother tongues. These are necessary to assess the potential influence of the mother tongue and consequently to produce the appropriate pedagogical aids for specific mother tongue groups. Comparisons between the different mother tongues and English will be made easier thanks to the bilingual computer corpora which are being collected today.

3) Good descriptions of learner use of prefabs. We need these descriptions as well as contrastive descriptions because not all learner problems are transfer-related. Computer learner corpora such as ICLE which cover different language backgrounds will make it possible to distinguish the phraseological features common to several categories of learners from the L1-dependent features.

4. Conclusion

Prefab-oriented approaches to teaching are currently in vogue, with EFL specialists suggesting that teaching procedures be based solidly on them. Yet when we consider how little we know about them, how they are acquired, what production difficulties they cause and how L1 and L2 prefabs interact, this is quite alarming. We possess insufficient knowledge to decide what role they should play in L2 teaching: we do not know what to teach, how much to teach and least of all how to teach, hence the urgent need for empirical work. This should be greatly facilitated by the wide variety of large computer corpora currently being assembled. The value of introspective tests in this field should also not be underestimated.

My own results indicate that the L1 plays an important role in the acquisition and use of prefabs in the L2. For obvious commercial reasons, most EFL material is aimed at all learners, irrespective of their mother tongue. Given the essentially language-specific nature of prefabs, this is a major issue that must be addressed if we are serious about giving learners the most efficient learning aids. Developing EFL materials from the types of data outlined above would go a long way towards solving this problem.

I gratefully acknowledge the support of the Belgian National Scientific Research Council and the University of Louvain Research Fund who helped to fund this research.
I wish to thank Carol Edgington for assisting me in collecting and analysing the data and Stephanie Tyson for helping me to clarify my vision of the final shape of the article.

Notes

1. The breakdown of the two corpora and their total number of words are given below:

Learner Corpus

ICLE subcorpus: French-speaking learners
164,190 words: untimed argumentative essays
24,174 words: timed argumentative essays
62,954 words: timed literature exam papers

TOTAL: 251,318 words

Native speaker corpus

1) Louvain essay corpus:
16,686 words: untimed argumentative essays
72,839 words: timed literature exam papers
2) International Corpus of English (ICE): (timed and untimed) student essays
50,202 words
3) LOB: Belles Lettres and essays (categories G36 - G77)
94,787 words

TOTAL: 234,514 words

Clearly, the size of the corpora used raises some questions for the study of prefabs. However, my research so far has demonstrated that, to quote S. JOHANSSON (1991), "there is still something to be said for the small, carefully-constructed corpus" and this is, in my view, especially true for learner language, which

is an extremely heterogeneous variety of English.

2. The following list is a selection of the booster combinations used exclusively by native speakers.

> acutely aware, astonishingly short, bitterly disillusioned, blatantly clear, blindingly obvious, brilliantly clever, devastatingly shocking, extensively excavated, extraordinarily painful, gravely disorganised, horribly disfigured, intensely aware, intimately bound up, irredeemably tied, irrevocably affected, keenly felt, ludicrously ineffective, mercilessly hard, monotonously uneventful, painfully clear, powerfully represented, profoundly shocked, readily available, ruthlessly callous, singularly stupid, steeply dipping, unbearably ugly, unusually small, vitally important.

3. In saying this, we do not disagree with I. MELCUK who pointed out at the symposium at which this paper was first presented that the use of **very** is not totally unrestricted and compared **very tired** and ***very rested** to demonstrate this. Nevertheless, **very** combines with more adjectives than any other amplifier and can, I think, therefore still correctly be termed the "all-round amplifier par excellence".

4. The 11 amplifiers presented were: highly, seriously, readily, blissfully, vitally, fully, perfectly, heavily, bitterly, absolutely, utterly. The format of the test was as follows:

readily significant reliable ill different essential aware
miserable available clear happy difficult ignorant
impossible cold important

---

bitterly significant reliable ill different essential aware
miserable available clear happy difficult ignorant
impossible cold important

5. B. ALTENBERG (this volume) refers to the high frequency of the epistemic stem **I think that** in spoken English. Arguably then, learners' overuse of this phrase may be a register-related problem.

6. C. LINDNER (1994) in his investigation of the German subcorpus of ICLE finds a similar overuse of **the fact that** in the English of advanced German EFL learners. His explanation for this is that "Apart from interference of German **die Tatsache, daß**, a missing flexibility on the part of the learners may also play a role. Their syntactic/phrasal repertoire when giving evidence for an observation is limited. Also, they may feel that expository-argumentative texts need a high degree of verbal factualness to be convincing".

This suggests that the overuse of this phrase and others may be partly due to transfer but also partly a common feature of learner writing.

7. There may well be individual differences here too. A. PETERS (1977, 571) suggests that there may be two types of adult L2 learner: the gestalt type, who prefers to learn a second language by feel, and the analytic type, who prefers to learn language 'by the book'.

Bibliography

AISENSTADT A. (1979). **Collocability restrictions in dictionaries**, in: R.R.K. HARTMANN (ed.) Dictionaries and their Users, Exeter Linguistic Studies, 71-74.

BAHNS J. (1993). **Lexical collocations: a contrastive view**. ELT Journal. Volume 47/1: 56-63.

BACKLUND U. (1973) **The Collocations of Adverbs of Degree in English**. Studia Anglistica Upsaliensia 13. Uppsala.

BENSON M., E. BENSON & R. ILSON (1986) **The BBI Combinatory Dictionary of English**. John Benjamins: Amsterdam & Philadelphia.

BOLINGER D.(1972) **Degree Words**. Mouton: The Hague & Paris.

CLEAR J.(1993) **From Firth Principles. Computational Tools for the Study of Collocation**, in: M. BAKER, G. FRANCIS & E. TOGNINI-BONELLI (eds.) Text and Technology. In Honour of John Sinclair. John Benjamins. Philadelphia & Amsterdam: 271- 292.

DECHERT H.(1984) **Second Language Production: Six Hypotheses**, in: H. DECHERT, D. MOHLE & M. RAUPACH (eds.) Second Language Productions. Gunter Narr Verlag. Tübingen: 211-230.

GRANGER S.(1993). **The International Corpus of Learner English**, in: J. AARTS, P. DE HAAN & N. OOSTDIJK (eds.) English Language Corpora: Design, Analysis and Exploitation. Rodopi. Amsterdam & Atlanta: 57-71.

GRANGER S.(1994). **The Learner Corpus: a Revolution in Applied Linguistics**. English Today

GRANGER S.(forthcoming) **A new kind of CA: Contrastive Interlanguage Analysis**, in: K. AIJMER & B. ALTENBERG (eds.) Languages in contrast. Text-based cross-linguistic studies, Lund Studies in English. Lund University Press.

GRANGER S.& S. TYSON (forthcoming). **Connector Usage in Native and Non-native English Essay Writing** (Revised version of a paper presented at the 10th AILA Congress, Amsterdam, 9-13 August 1993).

HANANIA E.& H. GRADMAN (1977). **Acquisition of English structures: a case study of an adult native speaker of Arabic in an English-speaking environment**. Language Learning. Volume 27/1: 75-91.

JOHANSSON S.(1991). **Times change, and so do corpora**, in: K. AIJMER & B. ALTENBERG (eds.) English Corpus Linguistics, Longman. London & New York: 305-314.

KJELLMER G.(1991). **A Mint of Phrases**, in: K. AIJMER & B. ALTENBERG (eds.) English Corpus Linguistics. Longman. London & New York: 111-127.

KJELLMER G.(1994). **A Dictionary of English Collocations: based on the Brown Corpus.** Clarendon Press. Oxford.

KRASHEN S.& R. SCARCELLA (1978). **On Routines and Patterns in Language Acquisition and Performance**. Language Learning. Volume 28/2: 283-300.

LEWIS M.(1993). **The Lexical Approach. The State of ELT and a Way Forward**. Language Teaching Publications.

LINDNER C.(1994). **Unnaturalness in Advanced Learners' English: A Corpus-Based Feasibility Study**. Unpublished MA Dissertation. Albert-Ludwigs-Universität Freiburg.

MOHLE D.& M. RAUPACH (1989). **Language Transfer of Procedural Knowledge**, in: H. DECHERT & M. RAUPACH (eds.) Transfer in Language Production. Ablex Publishing Corporation. Norwood, New Jersey: 195-216.

NATTINGER J.& J. DECARRICO (1992). **Lexical Phrases and Language Teaching**. Oxford University Press.

PETERS A.(1977). **Language learning strategies: does the whole equal the sum of the parts?**. Language. Volume 53/3: 560-573.

QUIRK R., S. GREENBAUM, G. LEECH & J. SVARTVIK (1985). **A Comprehensive Grammar of the English Language.** Longman. London & New York.

RAUPACH M.(1983). **Analysis and evaluation of communication strategies**, in: C. FAERCH & G. KASPER (eds.) Strategies in Interlanguage Communication. Longman. London & New York: 199-209.

RICHARDS J.(1983). **Communicative needs in foreign language learning**, ELT Journal. Volume 37/2: 111-120.

SINCLAIR J.(1987). **Collocation: a progress report**, in: R. STEELE & T. THREADGOLD (eds.) Language Topics. Essays in honour of Michael Halliday. Volume II. John Benjamins. Amsterdam & Philadelphia: 319-331.

VAN ROEY J.(1990) **French-English Contrastive Lexicology. An Introduction.** Louvain-la-Neuve. Peeters.

WILLIS D.(1990). **The Lexical Syllabus. A new approach to language teaching**. Collins ELT. London & Glasgow.

YORIO C.(1989) **Idiomaticity as an indicator of second language proficiency**, in: K. HYLTENSTAM & L.K. OBLER (eds.) Bilingualism across the Lifespan. Cambridge University Press: 55-72.