

Exemplification in learner writing: A cross-linguistic perspective

Magali Paquot

Centre for English Corpus Linguistics, Université catholique de Louvain – Belgium

[DRAFT]

Published in Meunier, F. & Granger, S. (eds) *Phraseology in Language Learning and Teaching*. Amsterdam: Benjamins.

1. Introduction

Recent corpus-based studies (e.g. Biber *et al.* 1999; Oakey 2002; Biber 2004; Biber *et al.* 2004) have pointed to the existence of an English for Academic Purposes (EAP) specific phraseology characterized by word combinations that are essentially semantically and syntactically compositional, e.g. *as a **result** of, in the **presence** of, the **aim** of this **study**, the **extent** to which, for **example**, it has been **suggested**, it should be **noted** that, it is **likely** that, as **shown** in **figure/fig.**, in **addition**, etc.* These word combinations are built around typical EAP or sub-technical words (in bold) and fulfil organizational or rhetorical functions that are prominent in academic writing, e.g. introducing a topic, hypothesizing, summarizing, contrasting, exemplifying, explaining, evaluating, concluding, etc.

Comparisons of native and learner corpora of academic writing have highlighted a number of features of non-nativeness or ‘unconventionality’ in the phraseology of learners of English as a Foreign Language (EFL). EFL learners have been shown to overuse a limited number of frequent English collocations and prefabs but to underuse a whole set of native-like phraseological units, especially typical EAP multi-word sequences, e.g. *claim that, the issue of, a strong argument* (De Cock 2003:364). Nesselhauf (2004:141) suggests that the unavailability of pragmatic chunks for the learners is most probably responsible for a number of deviant multi-word sequences which are used to structure the body of the essay, e.g. *Only have a look at; If you have a look at; Let us have a look at; A first argument I want to name for this*. The author also shows that among the nouns that are most often used with deviant verbs are typical EAP nouns like *action, aim, attitude, problem, question, statement, step* and *conclusion*.

Most of these studies have also pointed to the potential influence of the mother tongue on learners’ multi-word sequences. For example, Granger (1998) finds that the few English collocations involving intensifiers that are used by French learners typically have a direct

translation equivalent in French (e.g. *closely linked* 'étroitement lié'). Similarly, De Cock (2003) shows that French learners underuse a number of multi-word sequences which have no cognate forms in French (e.g. *sort of*), misuse some English sequences that have French partially deceptive cognates (e.g. *on the contrary* \approx 'au contraire', *in fact* \approx 'en fait') and use atypical combinations that are literal translations of French multi-word sequences (e.g. **according to me* 'selon moi').

Many studies mention the potential influence of the mother tongue on learners' production of multi-word sequences but very few, whether corpus-based or not, have tackled the issue systematically and examined the conditions under which multi-word sequences are most potentially transferable. In a number of studies based on acceptability tests and translation tasks, Kellerman (1977; 1978; 1979; 2000) suggests that L2 learners seem to work on the hypothesis that there are constraints on how similar the L2 can be to the L1, and these constraints seem to hold, even when the two languages are closely related and the structures congruent. Kellerman (1978) investigates the 'transferability' of the different meanings of the Dutch verb *breken* into its English cognate *break*. He shows that while Dutch learners of English accept the structures that are the least 'marked' in their mother tongue ('*he broke his leg*', '*the cup broke*'), they tend to reject what they perceive as 'language-specific' items ('*his voice broke when he was thirteen*', '*some workers broke the strike*'). 'Marked' in this context means "semantically odd, or syntactically less productible or less frequent when compared with 'normal' forms" (Kellerman 1979:46). In the 2000 study, Kellerman expands on these findings and argues that the dimension of 'prototypicality' largely determines Dutch learners' judgements about the transferability of the different usages of *breken* into *break*.

Although Kellerman acknowledges that learners' intuitions about what can be transferred in an L2 may not accurately reflect what they actually do when using the target language, his findings suggest that the further word combinations are situated from the central

core of phraseology, i.e. semantically opaque, syntactically and collocationally inflexible multi-word sequences, the more potentially transferable they may be. This conclusion is challenged by Nesselhauf in a study of learners' multi-word combinations with the two verbs *take* and *make* in which she claims that "it does not seem to be the case that transfer decreases with the degree of idiomaticity of a combination [...] but rather that locutional combinations [restricted collocations] – at least in the case of the verb-noun combinations with the two verbs investigated – are the type of combination that is most susceptible to transfer" (Nesselhauf 2003:278), e.g. **make part of* (Fr. 'faire partie de') for *be part of*, **make profit* (Fr. 'faire profit'), for *make a profit*, **make dreams* (Fr. 'faire des rêves') for *have dreams*. However, the author makes this claim on the basis of **erroneous** collocations only and does not examine the potential L1 influence on **native-like** multi-word sequences produced by learners.

The aim of this case study is to examine the potential influence of the mother tongue (L1) on learners' production of both correct and incorrect multi-word sequences that are typically used to fulfil an important rhetorical function, namely exemplification, in academic writing. In second language writing research, L1 influence has been shown to manifest itself in idea-generating and idea-organizing activities (Wang & Wen 2002). It may be suggested that EAP multi-word units are most potentially transferable not only because they are essentially semantically and syntactically compositional, i.e. typically unmarked word combinations, but also because they are directly anchored to an organizational or rhetorical function. The phraseological patterns of five exemplifying lexical items extracted from Paquot's (forthcoming) productively-oriented academic word list – the two fixed conjuncts *for example* and *for instance*, the noun *example* and the verbs *illustrate* and *exemplify* - are analyzed in five sub-corpora of the *International Corpus of Learner English* (Granger et al. 2002) with a view to distinguishing aspects of learner phraseological use characteristic of one

mother tongue background and therefore probably L1-dependent from phraseological patterns which are shared by most learner populations and hence more likely to be developmental or teaching-induced.

2. Data and methodology

This case study makes use of both native and learner corpora. The learner data consist of five sub-corpora of the *International Corpus of Learner English* (henceforth ICLE) of approximately 150,000 words each. Texts in each sub-corpus were carefully selected in an attempt to control external variables which may affect the written production of learners: they are all untimed argumentative essays written by higher-intermediate to advanced EFL university students of five different mother tongue backgrounds, namely Dutch, French, German, Polish and Spanish. An extended version of the *Louvain Corpus of Native Speaker Essays* (Granger 1996) (henceforth LOCNESS), a 300,000-word corpus of argumentative essays written by American university students, is used as a comparable corpus. Although numerous scholars have questioned the native speaker's status as the most relevant model for teaching English (e.g. Seidlhofer 2001; Jenkins 2005)¹, demonstrating a command of standard written English remains a high-priority requirement in academic settings (Flowerdew 2000; Hinkel 2004).

Jarvis's working definition of L1 influence refers to "any instance of learner data where a statistically significant correlation (or probability-based relation) is shown to exist between some feature of learners' IL [*interlanguage*] performance and their L1 background" (2000:252). The author suggests that to establish convincingly that an interlanguage behaviour exhibits L1-related effects, three potential effects of L1 influence should be systematically investigated, i.e. (1) intra-L1-group similarities, (2) inter-L1-group differences and (3) L1-interlanguage (IL) performance similarities. Intra-L1-group similarities are

similarities in the use of a common second language (L2) by learners from the same L1 background; inter-L1-group differences refer to differences in the use of a common L2 by learners from different mother tongue backgrounds; and L1-IL performance similarities are similarities in L1 and IL behaviour by learners who share the same mother tongue.

The *Integrated Contrastive Model* (Granger 1996; Gilquin 2000/2001) provides a very useful framework to investigate these three potential effects of L1 influence on the basis of naturally-occurring samples of learner language (Ellis & Barkhuizen 2005:25-30). The model combines Contrastive Analysis (CA) and *Contrastive Interlanguage Analysis* (CIA). CIA consists of two types of comparison. Firstly, it involves a comparison of native and non-native production of the same language which aims to “highlight a range of features of non-nativeness in learner writing and speech, i.e. not only errors but also instances of under- and overrepresentation of words, phrases and structures” (Granger 2002:12). Secondly, CIA includes a comparison of different interlanguages of the same language, i.e. the English of French learners, Spanish learners, Dutch learners, etc. This second type of comparison allows for the analysis of the first two types of L1-related effects: an investigation of a learner corpus made of texts written by learners of the same mother tongue background may reveal intra-L1-group similarities whereas a comparison of interlanguages by learners from at least two different L1 backgrounds may reflect inter-L1-group differences.

The combination of CA and CIA proposed by Granger (1996) helps analysts to link interlanguage behaviour to performance in the mother tongue and thus investigate the third type of potential effect of L1 influence, i.e. L1- interlanguage performance similarities. Similarities in L1 and IL behaviour by learners from the same mother tongue background are probably the strongest type of evidence as it “elucidates the relationship between the source and effects of L1 influence” (Jarvis 2000: 255). Corpora of texts written in the learners’ mother tongues are therefore used to investigate L1-IL performance similarities.

[Insert Figure 1]

As illustrated in Figure 1, The *Integrated Contrastive Model* proposes two types of approaches to CA and CIA data. From CA to CIA, the approach is predictive and consists in formulating CA-based predictions about L2 production which are then checked against CIA data. From CIA to CA, the approach is diagnostic: it aims to explain CIA findings, i.e. errors but also overuse and underuse, in the light of CA descriptions. The latter approach is adopted in this study but the more neutral term “explanatory” is preferred. Figure 2 shows that after a comparison of the learner corpora, i.e. different interlanguages, with the English native corpus (EN_{L1}) and subsequent comparisons between the learner corpora, results per L1 are paralleled with contrastive findings from L1 corpora when available.

[Insert Figure 2]

3. Exemplification in native writing

The lexical items that are most frequently used to fulfil an exemplifying function and that appear in Paquot’s (forthcoming) productively-oriented academic word list are of two types: (1) word sequences that intuitively appear to be single units and that will be called “word-like” units in this study, i.e. ‘*for example*’ and ‘*for instance*’ and (2) collocations and frames which do not intuitively appear to be single units as they are characterized by a high degree of variability but which are nevertheless repeatedly used to introduce examples, i.e.

collocations and frames with the noun ‘*example*’ and the verbs ‘*illustrate*’ and ‘*exemplify*’ (cf. Table 1).

[Insert Table 1]

‘*For example*’ is the most frequently used lexical item to introduce an example in native writing (104 occurrences in LOCNESS). It is typically used at the beginning of a sentence or after the subject:

1. *The arguments in favour of capital punishment revolve around the ancient biblical concept of “an eye for an eye”. **For example**, a criminal who murders should himself be murdered to fulfill what is considered by many to be justice. (LOCNESS)*
2. *Many of the fuels being developed today have little or no impact on the environment. **Hydrogen, for example**, burns completely clean. (LOCNESS)*

‘*For example*’ differs significantly from ‘*for instance*’ both in terms of frequency and register. Figure 3 shows the distribution of ‘*for example*’ and ‘*for instance*’ in the four sub-corpora of the BNC Baby, i.e. a four million word sampling of the 100 million word British National Corpus which represents the four main genres of academic writing, fiction, newspaper texts and spontaneous conversation.² A comparison of the four sub-corpora reveal that ‘*for instance*’ is less typical of academic prose. Moreover, ‘*for instance*’ is also generally much less frequent than ‘*for example*’ in English: it appears 187 times in the whole BNC Baby whereas ‘*for example*’ appears 832 times.

[Insert Figure3]

Productive frames with the noun ‘*example*’ and the verb ‘*to be*’, i.e. “X is DET (ADJ.) example of Y” and “DET (ADJ.) example of Y is X”, are also commonly found in academic writing:

3. *Non-source point pollution does not have a specific location and it unknowingly contaminates water supply. An example of this is leaking oil from an engine and the rain carrying it to a water source. (LOCNESS)*
4. *This is a prime example of thinking that does not follow in old footsteps but breaks away from convention and forges new routes.*
5. *Some of these diseases are life threatening. AIDS is a perfectly good example of these diseases. (LOCNESS)*

In LOCNESS, almost 50% of the occurrences of the noun ‘*example*’ outside the conjunct ‘*for example*’ appear in productive frames.

Finally, the verbs ‘*illustrate*’ and ‘*exemplify*’ are also used to introduce an example but ‘*illustrate*’ is much more frequent. They are typically found in the passive (cf. sentence 6) or in the active with non-human subjects such as *example, figure, table, case* and *approach* (cf. sentence 7).

6. *The mood of the time is illustrated by August Weismann who states: <*>. (LOCNESS)*
7. *As the above cases illustrate, the prayer that is proposed to be said in schools may have the opposite effect than what is intended. (LOCNESS)*

4. Exemplification in learner writing

An analysis of ‘*for example*’, ‘*for instance*’, ‘*example*’, ‘*illustrate*’ and ‘*exemplify*’ in learner writing shows that there are striking differences in use between native and learner writers of English. In section 4.1, patterns shared by the five learner populations are described while section 4.2 gives a description of features that are specific to the French learner population.

4.1. Shared patterns

Learners' use of word-like units displays similar patterns of overuse among the five mother tongue backgrounds: '*For example*' is overused by the five learner populations under investigation (cf. Figure 4) and '*for instance*' is massively overused by all learner populations except German learners (cf. Figure 5).³ Unlike in native writing, the two conjuncts are repeatedly found at the end of a sentence in learner writing:

8. *I find the pronunciation of English much more difficult than the pronunciation of Italian, **for instance**.* (ICLE-POLISH)
9. *Furthermore the psyche surpassed the nature of cosmic allegorism, which involved the representation of thunderstorm as a performance of the god Zeus **for example**.* (ICLE-DUTCH)

Non-native tendency to overuse the fixed formulae '*for example*' and '*for instance*' is clearly in line with Granger's (1998:156) conclusion that "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 in using."

[Insert Figure 4]

[Insert Figure 5]

It is most probable that learners 'cling on' to these two word-like units for two main reasons. Firstly, each first language has a direct equivalent to '*for example*', i.e. '*par exemple*' in French, '*por ejemplo*' in Spanish, '*bijvoorbeeld*' in Dutch, '*zum Beispiel*' in German and '*na przyktad*' in Polish, which provides a clear anchor in learners' minds. Secondly, word-like

units are typically focused on in instruction and teaching materials. Grammars and writing textbooks often only provide lists of adverbial connectors in sections on discursive functions. Similarly, instruction is most likely responsible for the massive overuse of ‘*for instance*’ in EFL students. ‘*For example*’ and ‘*for instance*’ are often taught as two synonyms and as a result, learners use them interchangeably. Learners presumably write ‘*for instance*’ instead of ‘*for example*’ when they have already used ‘*for example*’ in their essay and want to change for the sake of variety. However, it was shown in section 3 that ‘*for instance*’ is much less frequent than ‘*for example*’ in academic writing.

Another explanation for the general overuse of ‘*for example*’ and ‘*for instance*’ may also be that these word-like units are repeatedly used when they are unnecessary, redundant or even when other rhetorical functions should be made explicit. Sentence 10 is an example of redundant exemplification:

10. *The mob **for instance** is a very good **example**.* [ICLE-SPANISH]

The mirror image of this tendency is the underuse of the less salient multi-word units, i.e. productive frames with the noun ‘*example*’ and the verb ‘*to be*’, of which learners appear to be unaware. This underuse is also most probably reinforced by teaching as teaching materials very rarely put emphasis on these unmarked multi-word units. It is therefore not surprising that they are underused by almost all learner populations. The Dutch learners are the only ones who do not underuse these frames. This may be due to a difference in proficiency level as the Dutch sub-corpus represents the highest degree of proficiency in the ICLE corpus.⁴ Another possible explanation may be in terms of L1-related effects. In a comparable corpus of Dutch formal writing, productive frames with the verb ‘*zijn*’ were found to be quite frequent:

11. **Een goed voorbeeld is het akkoord dat eurocommissaris Bolkestein met de VS sloot over de verstrekking van passagiersgegevens van vluchten naar Amerika.**

12. **Een bekend voorbeeld is Linux, het alternatief voor het besturingssysteem van Microsoft.**

Finally, learners prove to underuse the verb ‘*exemplify*’. As to the verb ‘*illustrate*’, they underuse it only in the passive but overuse active structures with a human subject (13-14) and the phrase ‘*To illustrate this, ...*’ used at the beginning of a sentence (15-16):

13. ***I would like to illustrate that by means of some examples which, as you will see, are very diverse.*** (ICLE-DUTCH)

14. ***What I wanted to illustrate is that an obviously simple and clear solution turns out to be not the best one at a closer look – at least for one part of the people.*** (ICLE-GERMAN)

15. ***To illustrate the truth of this, one has only to mention people’s disappointment when realizing how little value has the time spent at university.*** (ICLE-SPANISH)

16. ***To illustrate this, we can mention the notion of culture and language in the north of Belgium.*** (ICLE-FRENCH)

4.2. Patterns specific to French learner writing

French learners repeatedly use the sequence *let’s take the example of*, a frame that is not incorrect but certainly very infrequent in native academic writing. Academic writing is a genre characterized by high degrees of formality and detachment and the speech-like nature of this sequence leads to an overall impression of stylistic inappropriateness:

17. *One form of this is nationalism. To show what I mean, let's take the example of an Englishman in Belgium.* (ICLE-FRENCH)

18. *One of them is the loss of contacts in families. (...) Let us take an example: many people eat while watching TV. I personally think that this is a pity.* (ICLE-FRENCH)

Whereas the conjunct ‘*for example*’ is significantly overused in most ICLE subcorpora, the frame *let's take the example of* is quite rare in other interlanguages and seems to be specific to French learners. This over-representation is most probably an L1-related effect as the sequence has a congruent counterpart in French, i.e. *Prenons l'exemple de ...*, which is commonly found in student and professional formal writing. The following sentences come from the *Corpus de Dissertations Françaises* (CODIF), a 200,000-word corpus of essays written by French-speaking students:

19. *Prenons l'exemple des sorciers ou des magiciens au Moyen Age.*

20. *Prenons l'exemple d'un individu qui postule pour un emploi.*

21. *Prenons l'exemple du port du voile qui a créé une polémique il y a tout juste 10 mois.*

In addition, Siepmann (2005:119-121) has shown that, contrary to English, French shows much greater reliance on the hortative to introduce an example, e.g. *considérons par exemple, citons l'exemple de, reprenons l'exemple de*, etc.

The overuse of ‘*let's take the example of*’ can be generalised to a massive overuse of ‘*let us / let's*’ in ICLE-FRENCH. Figure 6 shows that although ‘*let us*’ is generally more frequent in learner writing than in LOCNESS, the sequence is massively overused in the French learner corpus only.

[Insert Figure 6]

A few examples of sentences introduced with ‘*let us*’ in ICLE-FRENCH are:

22. *Let's consider the situation in Belgium.*
23. *Let's first have a look at what is Europe actually.*
24. *Now let's move on to our third category of criminals.*
25. *Let's try to find the most important principles which are urging people to react as they do.*
26. *So let us analyse the potential assets of this country...*
27. *Let us comment on the second statement: ...*
28. *Let us now examine the second solution.*
29. *Let us explain these two points.*

This massive overuse can be paralleled with a very frequent use of the first imperative plural to organise discourse in French formal writing:

30. *Citons comme exemple le jugement difficile des autorités françaises sur les activités du régime de Vichy dans le cadre du procès de Touvier. (CODIF)*
31. *Envisageons tout d'abord la question économique. (CODIF)*
32. *Examinons quelques exemples pour tenter d'y voir plus clair. (CODIF)*
33. *Et notons que ces réalisations nous inspirent des rêves insoupçonnés jusque là, ...*
34. *Considérons un instant le cinéma actuel. (CODIF)*
35. *Pensons, par exemple, à l'Espagne, qui, pendant quatre à huit siècles, a appris à côtoyer les peuples arabes. (CODIF)*

Furthermore, frequencies of first imperative plural forms in ICLE-FRENCH and CODIF are quite similar (cf. Table 2) and suggest that frequency in the learners' L1 may be reflected in their interlanguage.

[Insert Table 2]

This hypothesis is further supported by the similarity in frequency of first plural imperative forms in ICLE-Spanish and in the written part of the *Corpus de Referencia del Español Actual* (CREA).⁵ ‘*Let us*’ is very frequent in the Spanish learner corpus though not as frequent as in French learner writing. Similarly, Spanish first imperative plural forms are quite frequent in formal writing though less frequent than their French equivalents. The following sentences are examples of Spanish first imperative plural forms used as metadiscourse markers:

36. *Consideremos* ahora la distribución de la producción mundial de energía y de las emisiones de CO2 ... (CREA)

37. Pero *veamos* lo que se nos dice. (CREA)

38. Como ilustración, *tomemos* el caso de las alas de los pájaros. (CREA)

Finally, the infrequent use of ‘*laten we*’ in Dutch formal writing (1.4 occurrence per 100,000 word in the written part of the PAROLE corpus⁶) is also reflected in Dutch learner writing as ICLE-DUTCH is the only learner corpus in which the overuse of ‘let us’ is not statistically significant.

5. Transfer of L1 multi-word units

From this case study, there appear to be two types of transfer of L1 exemplifying multi-word units. The first type applies to word-like units and the second to less salient multi-word units. L2 word-like units tend to be overused by learners especially if they have similar units performing similar functions in their L1s. The overuse is often reinforced by instruction as teaching materials tend to focus exclusively on these units. This case study has shown that learners not only establish equivalence between ‘*for example*’ and the functionally equivalent word-like unit in their L1 (cf. Figure 7) but they also extend the equivalence to ‘*for instance*’.

One possible explanation for this rough equivalence is again the influence of instruction or ‘transfer of training’ (cf. Selinker 1972): although they differ in terms of frequency and register, ‘*for example*’ and ‘*for instance*’ are taught as functionally equivalent forms.

[Insert Figure 7]

The second type of transfer concerns more variable or less salient multi-word units that are essentially semantically and syntactically compositional, i.e. typically unmarked word combinations. French learners’ use of the multi-word unit ‘*let’s take the example of*’ is a direct translation of the French ‘*Prenons l’exemple de*’ but can be regarded as a case of positive transfer as the sequence is found in English. However, this multi-word unit in learner writing does not have the same stylistic profile as in native English writing. Figure 8 shows that transfer of form may not only go together with transfer of function but also with transfer of register and frequency. Thus, the L1 multi-word unit ‘*Let’s take the example of..*’ adopts the stylistic profile of the French sequence ‘*Prenons l’exemple de ..*’ and is repeatedly used in EFL French learner formal writing. An even more convincing example is the generalised overuse of the first person plural imperative in EFL French learner writing, a rhetorical strategy that does not respect English academic writing conventions but conforms to French academic style.

[Insert Figure 8]

6. Pedagogical implications

Descriptions of EAP vocabulary and its phraseology in native corpora are becoming available (cf. Verdaguer and González 2004) but information derived from learner corpus analysis, and more crucially, L1-specific information, is currently sorely lacking (Granger 2004). Patterns

shared by all learner populations can be used to increase the pedagogical value of teaching materials which would then incorporate information on aspects of overuse, underuse, misuse and learners' idiosyncratic use of lexico-grammatical means to fulfil rhetorical functions (cf. Flowerdew 1998). Learner-corpus analyses will enable the lexical means typically underused by all learner populations to be presented as useful alternatives to learners' preferred patterns. Thus, EAP textbooks should draw learners' attention to the fact that EFL students generally tend to overuse '*for example*' and '*for instance*' and advise them to also use productive frames such as '*X is a (adj.) example of Y*' and '*a (adj.) example of Y is X*' as well. Similarly, instead of presenting '*for example*' and '*for instance*' as interchangeable sequences, teaching materials should differentiate them and advise learners to make limited use of '*for instance*' in formal writing.

The findings presented in this study also support the view of contrastive rhetoric that "the linguistic patterns and rhetorical conventions of the L1 often transfer to writing in ESL and thus cause interference" (Connor 2002:494). They have important implications for EAP teaching as they suggest that the mother tongue deserves a place in the academic writing class. Although L1-specific findings may be more difficult to incorporate into generic EAP textbooks, these findings may prove particularly useful to EFL learners to help them conform to "the native stylistic norms for a particular register", which "entails not only making appropriate grammatical and lexical choices but also selecting conventional [multi-word units] to an appropriate extent" (Howarth 1998:186). It is highly desirable that they be incorporated into EAP textbooks specifically targeting EFL learners sharing the same mother tongue background or into electronic dictionaries and newly developed teaching materials such as writing aid software, where space is no longer so much of an issue (cf. Granger 2004).

7. Conclusion

The study shows that L1-related effects contribute significantly to learners' use of multi-word units in L2 together with other factors like transfer of training and level of proficiency.

Results suggest that transfer of form often goes together with transfer of function and supports the hypothesis that multi-word units with a pragmatic anchor are more easily transferred.

Transfer of form also seems to go together with transfer of frequency and register. These two types of transfer are less documented and well worth exploring in more detail. For learners, things can go wrong at each level: the **function** of a formally equivalent multi-word unit in the L2 is not always the same as in the L1; the **frequency** of a multi-word unit in L2 is often different from the frequency of its counterpart in L1 and in addition, there may be mismatches in **register** as shown by the example of the first person plural imperative in French vs. English academic writing.

Finally, this study has demonstrated the usefulness of the *Integrated Contrastive Model* to identify L1 influence. If we want to learn more about EAP multi-word units in learner writing and the conditions under which L1 sequences are most potentially transferable, we need three types of comparison: (1) comparisons of EAP vocabulary and its phraseology in learner vs. native writing, (2) comparisons of EAP vocabulary and its phraseology in different L1 learner corpora and (3) detailed descriptions of EAP vocabulary and its phraseology in other languages, such as the pioneering work of Siepmann (2005). It is only through a “constant to-ing and fro-ing between CA and CIA” (Granger 1996:46) that we can hope to gain a better understanding of the interaction between the L1 and L2 phrasicons.

Acknowledgements

I gratefully acknowledge the support of the Communauté française de Belgique, which funded this research within the framework of the ‘Action de recherche concertée’ project entitled ‘Foreign Language Learning: Phraseology and Discourse’ (No. 03/08-301). I would

also like to express my deep gratitude to Professor Sylviane Granger for her guidance and support.

Endnotes

1. These authors call for the development of a model based on proficient users of English as a Lingua Franca (ELF) on the basis that English is now used by many more non-native speakers than native speakers (cf. Modiano's (1999) Modified Concentric Circles). However, I share Mukherjee's (2005) doubts about "the suitability of this kind of English as a target norm for the ELT classroom."
2. For more information on the British National Corpus and the BNC Baby, cf. <http://www.natcorp.ox.ac.uk/>
3. Siepmann (2005) found that both '*for example*' and '*for instance*' were slightly less frequent in his non-native corpus of German student essays than they were in his native corpus. This difference in results is most probably due to a difference in text type as the author compares EFL student writing to professional writing.
4. A number of texts written by learners from the 11 mother tongue backgrounds found in the *International Corpus of Learner English* have recently been rated externally by a professional ex-ESOL rater according to the descriptors for writing found in the Common European Framework of Reference for Languages. Results show that learner essays rate from B2 to C2 and that the proportion of B2, C1 and C2 texts differs between the 11 mother tongue backgrounds (cf. Thewissen et al. 2006).
5. We make use of this corpus for lack of any comparable corpus of essays written by Spanish-speaking students. For more information on the *Corpus de Referencia del Español Actual*, see <http://www.rae.es>.
6. For more information on the PAROLE corpus, see <http://parole.inl.nl/html/index.html>

References

- Biber D. (2004) Lexical bundles in academic speech and writing. In Lewandowska-Tomaszczyk B. (ed.) *Practical Applications in Language and Computers* (PALC 2003). Frankfurt am Main: Peter Lang, 165-178.
- Biber D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999) *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Biber D., S. Conrad & V. Cortes (2004) *If you look at ...: Lexical bundles in university teaching and textbooks*. *Applied Linguistics* 25(3): 371-405.
- Connor U. (2002) New directions in contrastive rhetoric. *TESOL Quarterly* 36(4): 493-510.
- De Cock S. (2003) Recurrent Sequences of Words in Native Speaker and Advanced Learner Spoken and Written English: a Corpus-driven Approach. Unpublished PhD dissertation. Louvain-la-Neuve: Université catholique de Louvain.
- Ellis R. & G. Barkhuizen (2005) *Analysing Learner Language*. Oxford: Oxford University Press.
- Flowerdew L. (1998) Integrating 'Expert' and 'Interlanguage' Computer Corpora Findings on Causality: Discoveries for Teachers and Students. *English for Specific Purposes* 17(4): 329-345.
- Flowerdew J. (2000) Discourse Community, Legitimate Peripheral Participation, and the Nonnative-English-Speaking Scholar. *TESOL Quarterly* 34(1):127-149.
- Gilquin G. (2000/2001) The Integrated Contrastive Model: Spicing up your data. *Languages in Contrast* 3(1):95-123.
- Granger S. (1996) From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In Aijmer K., Altenberg B. & Johansson M. (eds) *Languages in Contrast. Text-based cross-linguistic studies. Lund Studies in English* 88. Lund: Lund University Press, 37-51. Available from <http://cecl.fltr.ucl.ac.be/>
- Granger S. (1998) Prefabricated patterns in advanced EFL writing: collocations and formulae. In Cowie A. (ed.) *Phraseology: theory, analysis and applications*. Oxford: OUP, 145-160.
- Granger S. (2002) A Bird's-eye view of learner corpus research. In Granger S., Hung J. & Petch-Tyson S. (eds) *Computer Learner Corpora, Second Language Acquisition and Foreign Language*

- Teaching. Language Learning and Language Teaching* 6. Amsterdam & Philadelphia: Benjamins, 3-33.
- Granger S. (2004) Computer learner corpus research: current status and future prospects. In Connor U. & T. Upton (eds) *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam & Atlanta: Rodopi, 123-145.
- Granger S., E. Dagneaux & F. Meunier (eds) (2002) *The International Corpus of Learner English. CD-ROM and Handbook*. Louvain-la-Neuve : Presses universitaires de Louvain. Available from <http://www.i6doc.com>
- Hinkel E. (2004) *Teaching Academic ESL Writing*. Mahwah, New Jersey & London: Lawrence Erlbaum Associates.
- Howarth P. (1998) The Phraseology of Learners' Academic Writing. In Cowie A.P. (ed.) *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press, 161-186.
- Jarvis S. (2000) Methodological rigor in the study of transfer: identifying L1 influence in the interlanguage lexicon. *Language Learning* 50(2): 245-309.
- Jenkins J. (2005) ELF at the gate: the position of English as a Lingua Franca. *Humanising Language Teaching* 7(2). Retrieved from www.hltmag.co.uk/mar05/idea.htm.
- Kellerman E. (1977) Towards a characterization of the strategy of transfer in second language learning. *Interlanguage Studies Bulletin* 2(1): 58-145.
- Kellerman E. (1978) Giving learners a break: native language intuitions as a source of predictions about transferability. *Working Papers on Bilingualism* 15: 59-92.
- Kellerman E. (1979) Transfer and non-transfer: where are we now? *Studies in Second Language Acquisition* 2: 37-57.
- Kellerman E. (2000) Lo que la fruta puede decirnos acerca de la transferencia léxico-semántica : una dimensión no estructural de las percepciones que tiene el aprendiz sobre las relaciones lingüísticas. In Muñoz C. (ed.) *Segundas lenguas: adquisición en el aula*. Barcelona: Ariel, 21-37.
- Modiano M. (1999) International English in the global village. *English Today* 15: 22-28.
- Mukherjee J. (2005) The Native Speaker is Alive and Kicking – Linguistic and Language Pedagogical Perspectives. *Anglistik* 16(2):7-23.

- Nesselhauf N. (2003) Transfer at the locutional level: an investigation of German-speaking and French-speaking learners of English. In Tschichold C. (ed.) *English Core Linguistics. Essays in honour of D.J. Allerton*. Bern: Peter Lang, 269-286.
- Nesselhauf N. (2004) *Collocations in a Learner Corpus*. Amsterdam: Benjamins.
- Oakey D. (2002) Formulaic language in English academic writing: A corpus-based study of the formal and functional variation of a lexical phrase in different academic disciplines. In Reppen R., S.M. Fitzmaurice & D. Biber (eds) *Using Corpora to Explore Linguistic Variation*. Amsterdam & Philadelphia: Longman, 111-129.
- Paquot M. (forthcoming) Towards a productively-oriented academic word list. In Walinski J., K. Kredens & S. Gozdz-Roszkowski (eds) *PALC 2005 Proceedings*. Frankfurt am Main: Peter Lang.
- Seidlhofer B. (2001) Closing a conceptual gap: the case for a description of English as a lingua franca. *International Journal of Applied Linguistics* 11(2):133-58.
- Selinker L. (1972) Interlanguage. *International Review of Applied Linguistics* 10(3):209-231.
- Siepmann D. (2005) *Discourse Markers Across Languages: A contrastive study of second-level discourse markers in native and non-native text with implications for general and pedagogic lexicography*. London: Routledge.
- Thewissen J., Y. Bestgen & S. Granger (2006) Using error-tagged learner corpora to create English-specific CEF descriptors. Paper presented at the Third Annual Conference of EALTA, 19-21 May 2006, Krakow, Poland.
- Verdaguer I. & E. González (2004) A Lexical Database of Collocations in Scientific English: Preliminary Considerations. In Williams G. & S. Vessier (eds) *Proceedings of the Eleventh EURALEX International Congress*. Lorient: Université de Bretagne-Sud, 929-934.
- Wang W. & Q. Wen (2002) L1 use in the L2 composing process: an exploratory study of 16 Chinese EFL writers. *Journal of Second Language Writing* 11:225-246.

Figure 1: The Integrated Contrastive Model (Granger 1996)

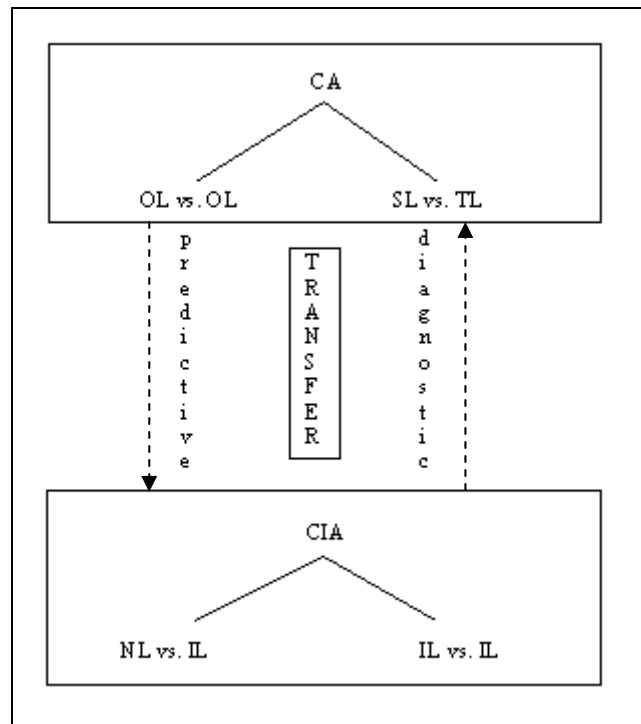


Figure 2: The Explanatory Version of the *Integrated Contrastive Model* (adapted from Granger 1996)

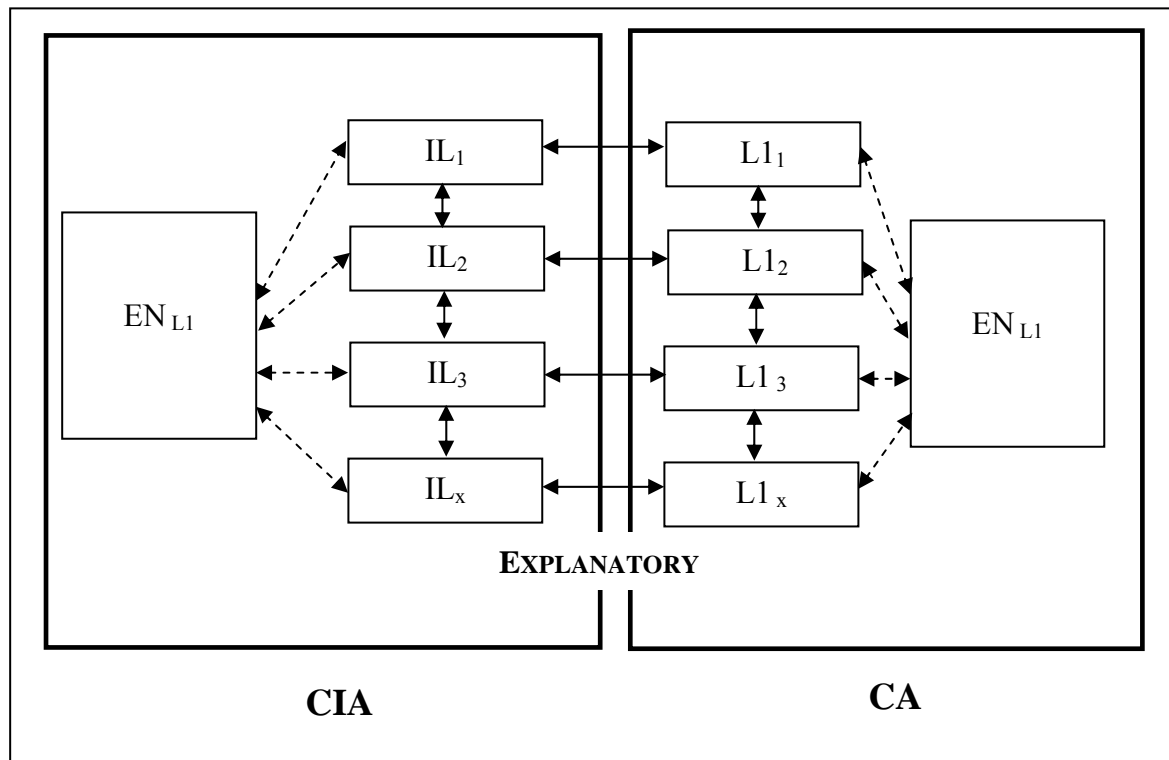


Table 1: Exemplifying function

Word-like units	Collocations and frames with...
For example	example
For instance	illustrate
	exemplify

Figure 3: Distribution of 'for example' and 'for instance'

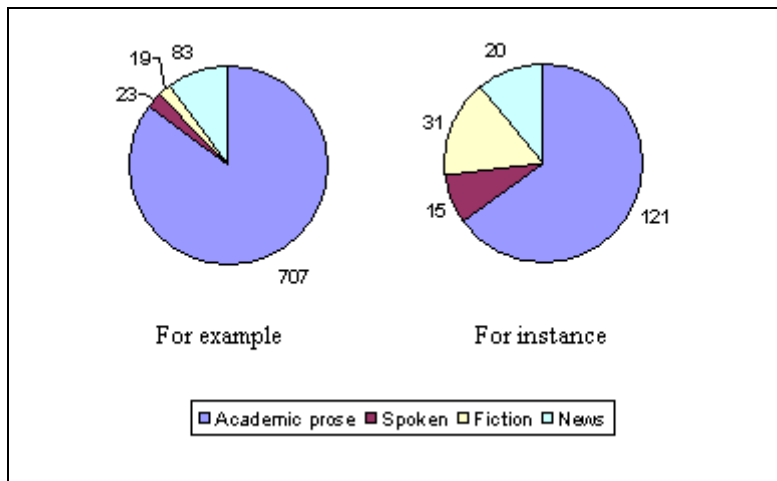


Figure 4: Number of occurrences of 'for example' per 100,000 words

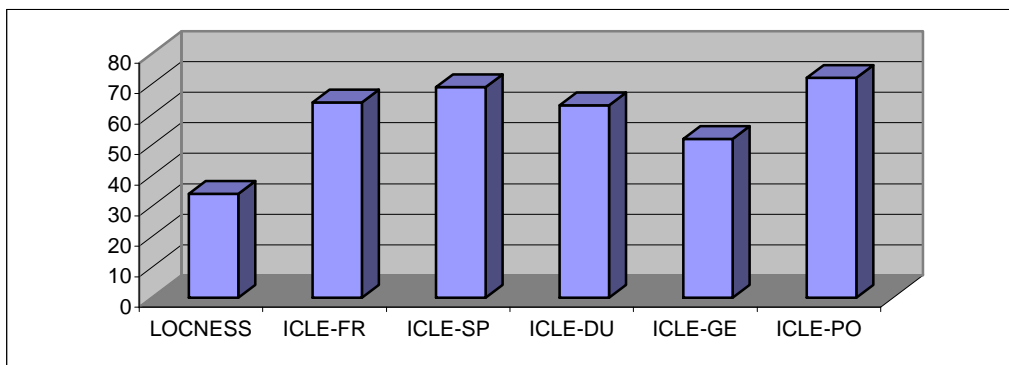


Figure 5: Number of occurrences of 'for instance' per 100,000 words

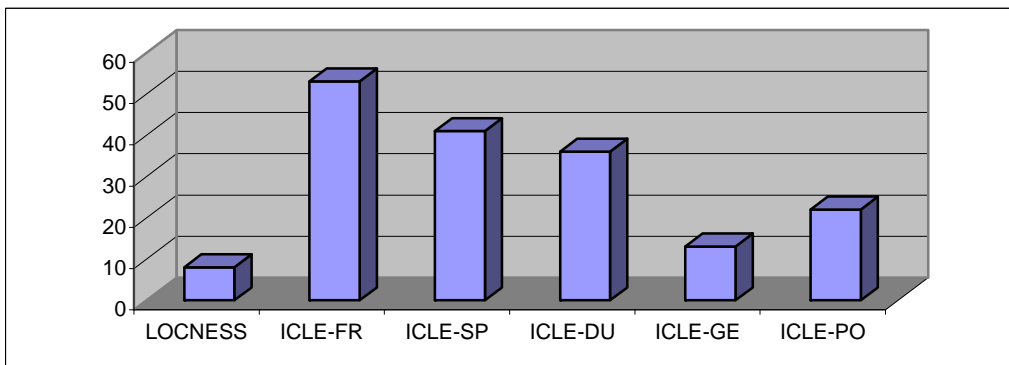


Figure 6: Number of occurrences of 'let us' and 'let's' per 100,000 words

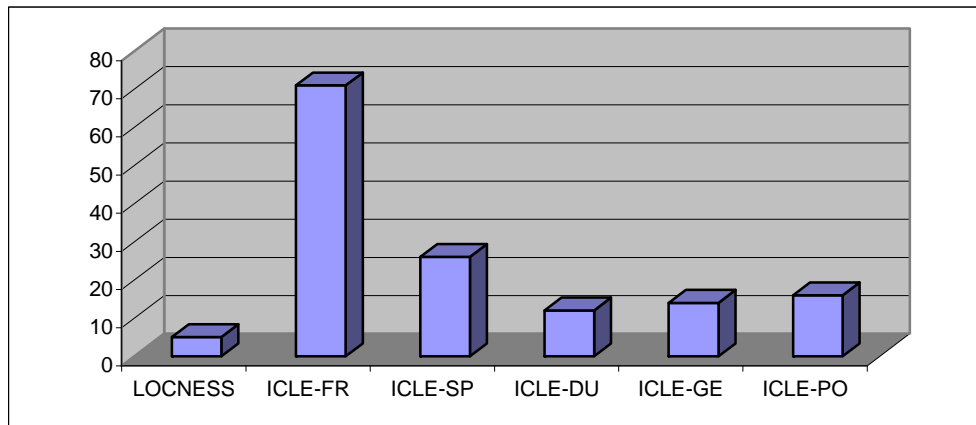


Table 2: First imperative plural forms in ICLE-FRENCH and FREDA

Corpora	Relative frequencies per 100,000 words
ICLE-FRENCH	71
FREDA	85

Figure 7: Transfer of L1 word-like units

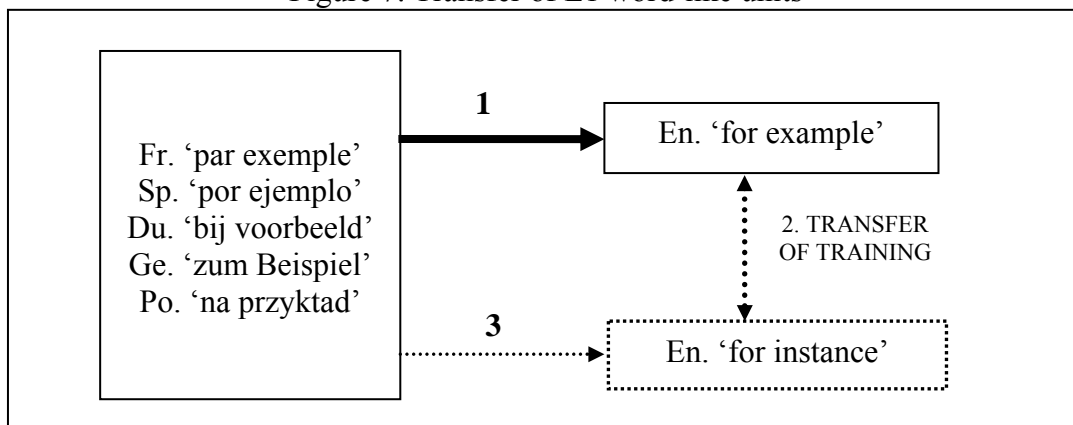


Figure 8: Transfer of L1 less salient multi-word units

