

EAP vocabulary in learner corpora

A cross-linguistic perspective

Magali Paquot

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

Keywords: English for Academic Purposes (EAP), EAP vocabulary, EAP clusters, learner corpora, L1 influence

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 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. These findings support Howarth's claim that it is not idioms that learners need for effective communication in academic settings but the lexical means that will allow them to 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).¹

2. The phraseology of EFL academic writing

Comparisons of native and learner corpora of academic writing have highlighted a number of features of non-nativeness or 'unconventionality' in the phraseology of EFL students. Learners have been shown to overuse a limited number of frequent English collocations and prefabs (Flowerdew 1998, 2003; Kaszubski 2000; De Cock 2003, this volume) but to underuse a whole set of native-like collocations, 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 [...] appears to be the underlying reason for a number of deviant collocations 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 underlined learners' "underdeveloped sense of salience of what constitutes a significant collocation" (Granger 1998:152) and pointed to the potential

¹ Although numerous scholars have questioned the native speaker's status as the most relevant model for teaching English, the fact remains that demonstrating a command of standard written English is a high-priority requirement in academic settings (Hinkel 2004).

influence of the mother tongue on learners' multi-word sequences. For example, Granger (1998) found 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) showed 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* ≈ 'au contraire', *in fact* ≈ 'en fait') and use atypical combinations that are literal translations of French multi-word sequences (e.g. **according to me* 'selon moi').

3. The potential influence of the mother tongue on learners' multi-word sequences

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) suggested 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) investigated the 'transferability' of the different meanings of the Dutch verb *breken* into its English cognate *break*. He showed that while Dutch learners of English accepted the structures that were the least 'marked' in their mother tongue ('*he broke his leg*', '*the cup broke*'), they tended to reject what they perceived 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 expanded on these findings and argued that the dimension of 'prototypicality' largely determined Dutch learners' judgements about the transferability of the different usages of *breken* into *break*.

Although Kellerman acknowledged 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 (cf. Granger: this volume), the more potentially transferable they may be. This conclusion is nevertheless 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 on learners' production of both correct and incorrect EAP multi-word sequences. In contrastive rhetoric, L1 influence has been shown to manifest itself "in the writer's choice of rhetorical strategies and content" (Connor 2002:494). It may be suggested that EAP multi-word sequences are most potentially transferable not only because they are essentially semantically

and syntactically compositional, i.e. typically unmarked word combinations (cf. section 1) but also because there are many congruent multi-word sequences in different languages that are directly anchored to a rhetorical function, such as *for example*, *par exemple* and *por ejemplo* - for the function of 'illustration'.

To test this hypothesis, the study makes use of both native and learner written data. The learner data consist of 5 sub-corpora of the *International Corpus of Learner English* (Granger *et al.* 2002) (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 (cf. Jarvis 2000): 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* (henceforth LOCNESS), a 300,000-word corpus of argumentative essays written by American university students, is used as a comparable corpus.

This case study examines the lexical means used by the five learner populations described above to serve an important rhetorical function in academic writing, namely exemplification. The phraseological patterns of four exemplifying words extracted from Paquot's (2005) productively-oriented academic word list – *example*, *instance*, *exemplify*, *illustrate* - are analyzed with a view to assessing the potential influence of the mother tongue on learner use.

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.
- 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 L. (2003) A Combined Corpus and Systemic-Functional Analysis of the Problem-Solution Pattern in a Student and Professional Corpus of Technical Writing. *TESOL Quarterly* 37(3): 489-511.
- 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., 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.

- Kaszubski P. (2000) *Selected aspects of lexicon, phraseology and style in the writing of Polish advanced learners of English: a contrastive, corpus-based*. PhD Thesis. Poznań: Adam Mickiewicz University.
- 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.
- Jarvis S. (2000) Methodological rigor in the study of transfer: identifying L1 influence in the interlanguage lexicon. *Language Learning* 50(2): 245-309.
- 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. (2005) *Towards a productively-oriented academic word list*. Paper presented at Practical Applications in Language and Computers, 7-9 April 2005, Łódź, Poland.