The *Louvain EAP Dictionary (LEAD)*

Sylviane Granger and Magali Paquot[1]
Centre for English Corpus Linguistics, Université catholique de Louvain

In our software demonstration, we describe a web-based English for Academic Purposes dictionary-cum-writing aid tool, the Louvain EAP Dictionary (LEAD). The dictionary is based on the analysis of c. 900 academic words and phrases in a large corpus of academic texts and EFL learner corpora representing a wide range of L1 populations. The dictionary contains a rich description of non-technical academic words, with particular focus on their phraseology (collocations and recurrent phrases). Its main originality is its customisability: the content is automatically adapted to users' needs in terms of discipline and mother tongue background. Another key feature of the LEAD is that is makes full use of the capabilities afforded by the electronic medium in terms of multiplicity of access modes (Tarp 2009). The dictionary can be used as both a semasiological dictionary (from lexeme to meaning) and an onomasiological dictionary (from meaning/concept to lexeme) via a list of typical rhetorical or organisational functions in academic discourse (cf. Pecman 2008). It is also a semi-bilingual dictionary (cf. Laufer & Levitzky-Aviad 2006) as users who have selected a particular mother tongue background can search lexical entries via their translations into that language.

The LEAD is designed as an integrated tool where the actual dictionary part is linked up to other language resources and learning tools. It is a hybrid dictionary (cf. Hartman 2005) that includes both a dictionary-cum-corpus and a dictionary-cum-CALL component. As regards direct corpus access, the LEAD innovates by giving access to discipline-specific corpora rather than generic corpora.

While the current version of the tool is restricted to some disciplines and mother tongue backgrounds, its flexible architecture allows for further customisation (other L1 background populations, other disciplines, other languages).

1. Introduction

Recent research on written academic skills has considerably improved our understanding of the challenges faced by non-native speakers when they write academic texts in English. In particular, it has uncovered the role played by non-technical academic words to express key rhetorical or organisational functions such as contrasting, exemplifying or concluding (Thurstun and Candlin 1998). Corpus-based analyses have demonstrated a high degree of commonality in the use of these words by expert writers from different disciplines but have also highlighted a number of discipline-specific patterns that need to be described. At the same time, learner corpus research has identified the particular types of difficulty that these words pose to non-native writers and demonstrated the important role played by transfer from the learner's mother tongue (Paquot 2010).

In this article, we describe an attempt to implement these findings in a customisable web-based tool that aims to help learners produce better academic texts. Section 2 introduces the *Louvain EAP dictionary (LEAD)* and focuses on two of its major characteristics, i.e. its customisability and multiplicity of access modes. The *LEAD* is designed as an integrated tool where the actual dictionary part is linked up to other language resources and learning tools. In section 3, we describe how we are currently turning the *LEAD* into a hybrid dictionary (cf. Hartman, 2005) that includes both a dictionary-cum-corpus and a dictionary-cum-CALL component. Section 4 concludes.

[1] We gratefully acknowledge the support of the Belgian National Fund for Scientific Research (FNRS), which funded this research within the framework of the FRFC project 'Lexicography and phraseology: onomasiological and semasiological approach to English for Academic Purposes' (2.4.501.08.F).

2. The Louvain EAP Dictionary

The *Louvain EAP Dictionary* is a customisable web-based English for Academic Purposes (EAP) dictionary-cum-writing-aid tool (Granger and Paquot, in press). It is based on the analysis of c. 900 academic words and phrases in a large corpus of academic texts (i.e. the academic component of the British National Corpus as well as home-made discipline-specific corpora) and EFL learner corpora representing a wide range of L1 populations. The dictionary contains a rich description of the EAP words, with particular focus on their phraseology (collocations and recurrent phrases). Its main originality is its customisability: the content is automatically adapted to users' needs in terms of *discipline* and *mother tongue background*. Its underlying principles are close to Tarp's (2008; 2009) 'function theory', which sees lexicographic needs as 'related to specific types of users who find themselves in a specific type of social situation which, by definition, is extra-lexicographic' (Tarp 2009: 25). It relies on a relational MySQL database (cf. Verlinde et al. 2009), the technical characteristics of which make it possible to exploit linguistic information as a 'multifunctional lexicographical database', i.e. a 'modularly designed dictionary database targeting several kinds of users in many different user situations' (Pajzs 2009: 326).

A key feature of the *LEAD* is that is makes full use of the capabilities afforded by the electronic medium in terms of multiplicity of access modes (see Sobkowiac 2002; Tarp 2009). The dictionary can be used as both a *semasiological* dictionary (from lexeme to meaning) and an *onomasiological* dictionary (from meaning/concept to lexeme) via a list of typical rhetorical or organisational functions in academic discourse (cf. Pecman 2008). It is also a *semi-bilingual dictionary* (cf. Laufer & Levitzky-Aviad 2006) as users who have selected a particular mother tongue background can search lexical entries via their translations into that language.

Before using the dictionary, users have to select a domain (currently business, medicine, linguistics, or general EAP for users working in other disciplines) and specify their L1 background (currently French) (cf. Figure 1).

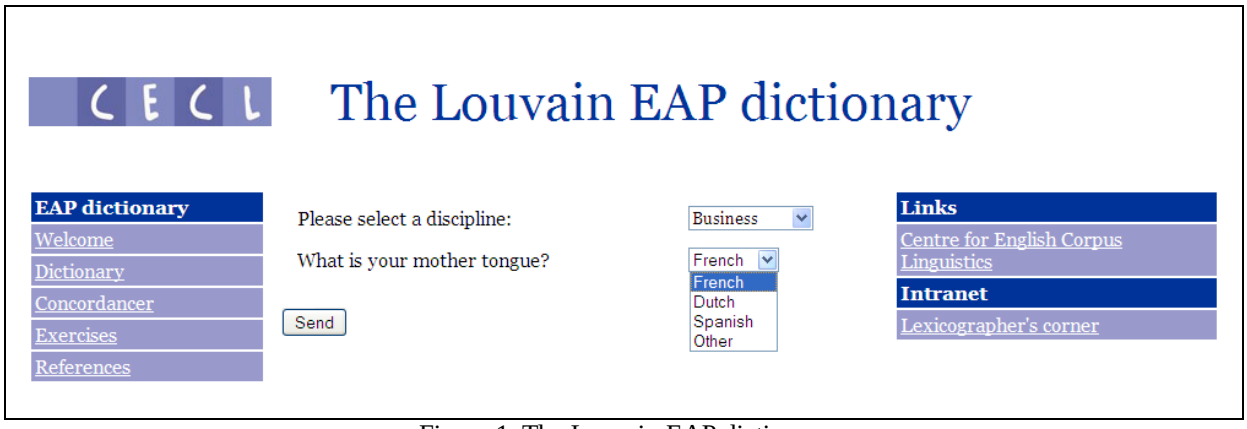

Figure 1. The Louvain EAP dictionary

Domain selection makes it possible to customise the output and illustrate the phraseological environment of a search word by means of example sentences extracted from a corpus of either business, medicine or linguistics texts. The characteristics of good dictionary examples have been clearly identified by Atkins & Rundell (2008: 458): they should be (1) natural and typical, (2) informative, and (3) intelligible. However, these are not intrinsic properties and they need to be customised to the type of dictionary and the needs of its users. In the *LEAD*,

the collocation *cause + distress* will therefore be illustrated by example (1) when the user has selected business as the target discipline and by example (2) when medicine is the target, thereby responding to Moon's (2008: 333) recommendation that particular attention be paid to 'the function of phraseological information in relation to the needs and interests of the target users' (Moon 2008: 333).

(1) *Rivals may not be able to bear initial losses, which would* ***cause financial distress*** *rather than lead to balanced growth.*
(2) *Severe hypoglycaemic attacks* ***cause distress*** *for diabetics and their families.*

One of the purposes of L1-background identification is to give *feedback* on errors and problems that a specific L1 population typically encounters. When the dictionary is used as a semi-bilingual dictionary, we also include warnings about common translation mistakes such as the erroneous translation of the French verb 'prétendre' by its false friend 'pretend' in English. We are currently focusing on French as a L1 background but are planning to include more languages in the future. To create both the generic usage notes and the L1-specific notes, we make use of the *International Corpus of Learner English* (Granger et al. 2009) as well as the *Varieties of English for Specific Purposes dAtabase* (VESPA), a new learner corpus, currently being developed in Louvain in collaboration with several international partners. The corpus contains L2 texts from a wide range of disciplines (linguistics, business, engineering, sociology, etc), genres (papers, reports, MA dissertations) and degrees of writer expertise in academic settings (from first-year students to PhD students) (see http://cecl.fltr.ucl.ac.be/VESPA.html for further details).

The *onomasiological* dictionary is based on a list of 18 rhetorical functions that we have identified as being particularly prominent in academic discourse, e.g. comparing and contrasting, expressing cause and effect, introducing a concession (Figure 2).

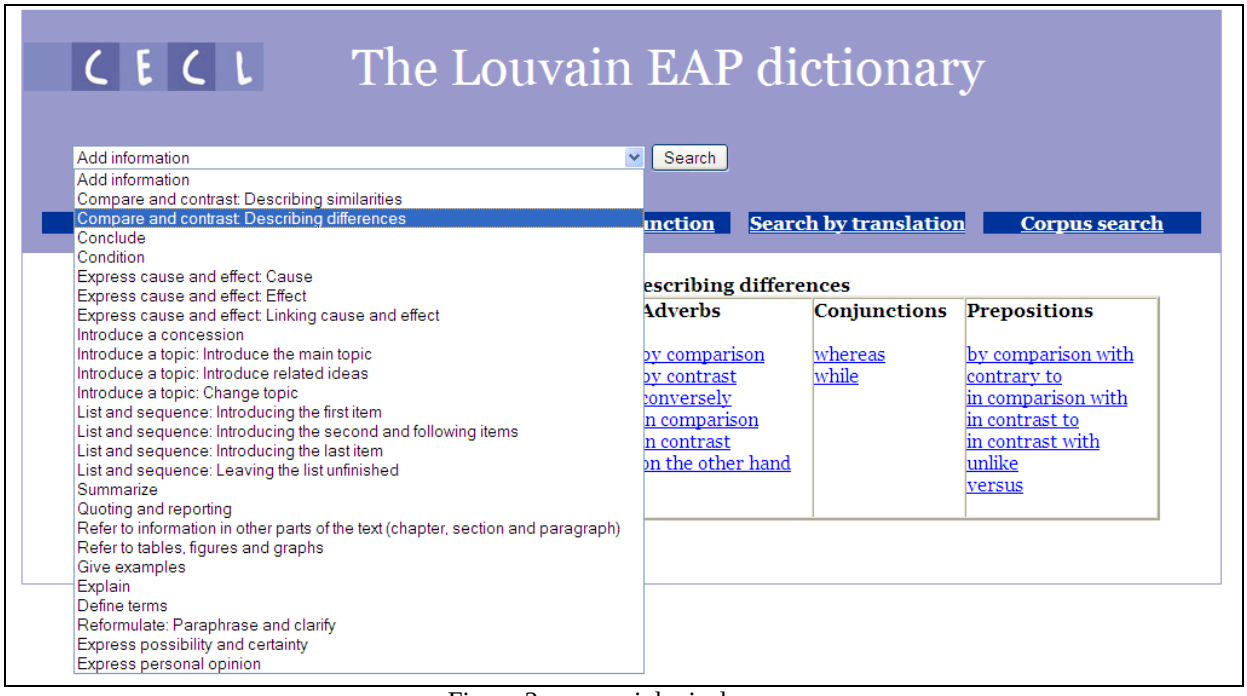

Figure 2. onomasiological access

Selecting one of these functions provides the user with a list of lexical items including nouns, verbs, adjectives, adverbs, and phrases that are typically used to serve this function in

academic texts. With this access mode, we hope to help users enlarge their lexical repertoire. Words are currently sorted alphabetically but in the future they could be sorted by frequency of occurrence in the discipline-specific corpora. Each word is clickable and users can get access to its lexical entry.

3. *LEAD*: a dictionary-cum-corpus and a dictionary-cum-CALL

To further help learners to master academic vocabulary, we are currently integrating the *LEAD* into a broader working environment that includes a corpus handling tool and exercises targeting the words and phrases included in the dictionary database.

As regards direct corpus access, the *LEAD* innovates by giving access to *discipline-specific corpora* rather than generic corpora. Discipline-specific corpora are accessible via the 'Corpus search' icon (cf. top right of Figure 2). We make use of a new open source web-based corpus analysis system, viz. CQPweb (http://sourceforge.net/projects/cwb/files/), developed by A. Hardie (Hardie, 2009) for two main purposes:

- give access to concordances of academic words and their collocations to provide users with more examples and make it possible for them to check whether a collocation or phrase that is not in the *LEAD* is correct or not;
- query words that are not in the *LEAD* so that users can check how to use a word even though it is not in the dictionary. As we are focusing on a very specific and quite limited vocabulary, we want to be able to provide another kind of feeback than the very frustrating 'No match found!' when the search word is not in the dictionary.

Abel (in press) uses the term 'dictionary-cum-CALL' to refer to electronic dictionaries that serve as the basis for the development of CALL modules (e.g. Verlinde et al 2007). As such, they constitute yet another type of mixed dictionary genres that are the result of 'hybridisation' with one or more types of reference work (Hartmann, 2005). We are currently turning the *LEAD* into a dictionary-cum-CALL resource by adding exercises targeting learners' attested difficulties. We make use of the Hot Potatoes software (http://hotpot.uvic.ca/index.php) to create exercises such as fill-in-the-blanks, word building, error detection and collocation exercises. In the same way as the dictionary is customisable to users' needs, our aim is to develop an adaptable learning environment that directs users to exercises targeting their specific needs. Thus, a French-speaking user will have access to a number of exercises that deal with French learners' specific difficulties (e.g. overuse of 'in fact' as a direct translation of 'en fait', heavy reliance on 'let us' to introduce examples, change topic or direct the reader's attention to specific facts). Exercises include error detection exercises that focus on L1-induced errors and translation quizzes in which the user must choose between several alternative English translations of a French sentence.

4. Conclusion

As English is incontestably the dominant language in academia, acquiring good English academic skills is mandatory for the large proportion of users for whom English is a non-native language. More and more university students have to write term papers, reports, or dissertations in English. The number of master and doctoral programmes taught in English has increased dramatically over the last decade. For researchers, the stakes are even higher as inappropriate language use is a major factor in the rejection of articles submitted to international journals by non-native writers (cf. Mungra & Weber 2010). The objective of the

LEAD is to meet the growing needs of non-native speakers, be they students of English or researchers, who have to write academic texts that conform to the established conventions of the genre (more particularly its phraseology). The automatic customisation of the dictionary to users' discipline and L1 coupled with direct corpus access makes it a particularly dynamic tool. The inclusion of error warnings and targeted exercises gives it the status of a hybrid tool, i.e. both a dictionary and a learning resource. The *LEAD* is also a highly flexible tool, which could easily be customised to other L1 background populations, other disciplines, and other languages.

References

Abel, A. (in press). 'Towards a systematic classification framework for dictionaries and CALL'. In Granger, S.; Paquot, M. (eds.) *eLexicography in the 21st century: New challenges, new applications. Proceedings of ELEX2009*. Louvain-la-Neuve, Presses universitaires de Louvain.

Atkins, S.; Rundell, M. (2008). *The Oxford Guide to Practical Lexicography.* Oxford: Oxford University Press.

Granger, S.; Dagneaux, E.; Meunier, F; Paquot, M. (2009). *The International Corpus of Learner English. Handbook and CD-ROM. Version 2.* Louvain-la-Neuve: Presses universitaires de Louvain.

Granger, S.; Paquot, M. (in press) 'Customising a General EAP Dictionary to Learner Needs'. In Granger, S.; Paquot, M. (eds.). *eLexicography in the 21st century: New challenges, new applications. Proceedings of ELEX2009*. Louvain-la-Neuve, Presses universitaires de Louvain.

Hardie, A. (2009). 'CQPweb – combining power, flexibility and usability in a corpus analysis tool'. Paper presented at the 30th ICAME conference, Lancaster, 27-31 May 2009.

Hartmann, R. R. K. (2005) 'Pure or hybrid? The development of mixed dictionary genres'. In *Facta Universitatis* 3 (2). 193-208. http://facta.junis.ni.ac.rs/lal/lal2005/lal2005-06.pdf [access date: 18 February 2010]

Laufer, B.; Levitzky-Aviad, T. (2006). 'Examining the effectiveness of of 'bilingual dictionary plus' – a dictionary for production in a foreign language'. In *International Journal of Lexicography* 19 (2). 135-155.

Moon, R. (2008). 'Dictionaries and collocation'. In Granger, S.; Meunier, F. (eds.), *Phraseology. An Interdisciplinary Perspective*. Amsterdam: Benjamins. 313-336.

Mungra, P.; Webber, P. (2010). 'Peer review process in medical research publications. Language and content comments'. In *English for Specific Purposes* 29 (1). 43-53.

Pajzs, J. (2009). 'On the possibility of creating multifunctional lexicographical databases'. In Bergenholtz, H.; Nielsen, S.; Tarp, S. (eds.). *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Bern: Peter Lang. 327-354.

Paquot, M. (2010). *Academic Vocabulary in Learner Writing*. London & New York: Continuum.

Pecman, M. (2008). 'Compilation, formalisation and presentation of bilingual phraseology: problems and possible solutions'. In Meunier, F.; Granger, S. (eds.), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam & Philadelphia: Benjamins. 203-222.

Sobkowiac, W. (2002). *What can be but is not (and why) in learners' MRDs. Teaching English with Technology* 2 (3). http://www.iatefl.org.pl/call/j_article9.htm [access date: 18 February 2010]

Tarp, S. (2008). *Lexicography in the Borderland between Knowledge and Non-Knowledge.* Tübingen: Niemeyer.

Tarp, S. (2009). 'Beyond lexicography: new visions and challenges in the information age'. In Bergenholtz, H.; Nielsen, S.; Tarp, S. (eds.). *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Bern: Peter Lang. 17-32.

Thurstun J.; Candlin, C. (1998). 'Concordancing and the teaching of the vocabulary of Academic English'. In *English for Specific Purposes* 17 (3). 267-280.

Verlinde, S.; Binon, J.; Ostyn, S.; Bertels, A. (2007). 'La Base lexicale du français (BLF): un portail pour l'apprentissage du lexique français'. In *Cahiers de lexicologie* 91 (2). 251-266.

Verlinde, S.; Selva, T.; Binon, J. (2009). 'Les bases de données au service d'un dictionnaire d'(auto-)apprentissage pour allophones'. In Corbin, P.; Gasiglia, N. (eds.). *Changer les dictionnaires ?* Paris: Presses universitaires du Septentrion. 217-234.