

Electronic lexicography goes local Design and structures of a needs-driven online academic writing aid

Sylviane Granger & Magali Paquot

Granger, S. & Paquot, M. (2015). Electronic lexicography goes local: Design and structures of a needs-driven online academic writing aid. *Lexicographica - International Annual for Lexicography / Internationales Jahrbuch für Lexikographie* 31(1), 118-141.

1. Introduction
2. General design of the *Louvain English for Academic Purposes Dictionary*
 - 2.1. Corpus-based
 - 2.2. Production-oriented
 - 2.3. Hybrid
 - 2.4. Customizable
3. Macrostructure
4. Microstructure
 - 4.1. The phraseological patterning of EAP words
 - 4.1.1. Collocations
 - 4.1.2. Lexical bundles
 - 4.1.3. Examples
 - 4.2. Other types of information
5. Access structure
6. Conclusion and perspectives

Abstract

Writing idiomatic academic texts in English is a challenge faced by an increasing number of international students and scholars. In this article, we present a web-based dictionary-cum-writing aid tool, the *Louvain English for Academic Purposes Dictionary*, which aims to address the attested needs of non-native writers of English, with a particular focus on the phraseological patterning of academic words. First, we present the general design of the dictionary, highlighting four of its main features, i.e. the fact that it is corpus-based, production-oriented, hybrid and customizable. We then describe key aspects of the macro- and microstructure and highlight innovative features of the access structure, such as direct access to discipline-specific corpora. The conclusion outlines some important priorities for the future, in particular the need for validation and the desirability of integrating the dictionary into a wider learning environment.

1. Introduction

A good command of English for academic purposes (EAP) is of critical importance for an ever-increasing number of students and researchers who are non-native speakers of English. There has been a steep growth in the number of English-taught Master's and doctoral degree programmes internationally, all of which involve written assignments in English. The need is even more crucial for researchers, as English is the dominant language in most of the top international journals, to the extent that the "publish or perish" imperative which governs a researcher's career currently reads as "publish in English or perish". Fortunately, recent

developments in the field of specialized lexicography are going some way towards meeting non-native users' needs. A new type of dictionary, the "specialized dictionary for learners" (Fuertes-Olivera 2010), caters for learners' needs with regard to discipline-specific words such as *firewall* (computing) or *liposome* (medicine). While in the past "the main purpose of monolingual specialised dictionaries tended to be to clarify meaning rather than specify usage" (Pearson 1998: 71), the new specialized dictionaries for learners also provide comprehensive information on "the linguistic functioning of terms (examples, explanations about subtle semantic distinctions, collocates, syntactic structure, etc.)" (L'Homme 2010: 141). However, general academic words – which, unlike specialized content words, cut across disciplines – are largely unaffected by these developments. Words such as *cause*, *therefore*, *argue* and *significant*, which are used to "refer to those activities that characterize academic work, organize scientific discourse and build the rhetoric of academic texts" (Paquot 2010: 28), are usually absent from specialized dictionaries.

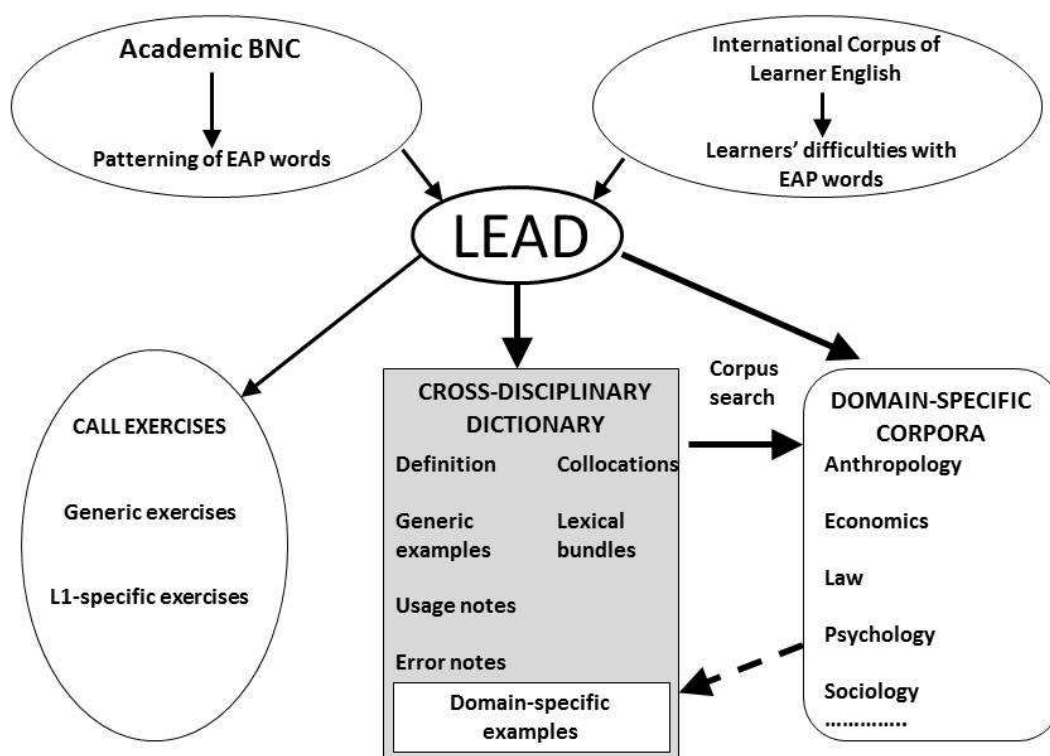
The pervasive neglect of general academic words is problematic, as these words have their own EAP-specific patterning of which non-native speakers of English (and indeed many native speakers) are largely unaware. This distinctive lexico-grammatical patterning has been clearly brought out by corpus-based studies of academic texts. For example, Flowerdew (2008) identified the typical patterning of words expressing 'problem and solution' (e.g. *problem*, *need*, *solution*, *propose*, *recommend*) in a professional academic corpus and compared it with their patterning in a corpus of texts written by Hong Kong undergraduates. She concludes that student use, though generally grammatical, is often unidiomatic because students "lack the necessary lexico-grammar for expressing their ideas" (p. 132).

A large number of studies based on learner corpora has generated results that are in line with Flowerdew's conclusion (for a survey, see Paquot & Granger 2012). This prompted us to embark on the design and implementation of a production-oriented online tool, the *Louvain English for Academic Purposes Dictionary* (LEAD), which would cater for learners' general academic needs. The design of the resource is firmly rooted in British pedagogical lexicography, which has played a key role in turning dictionaries into efficient production tools. One of the distinctive features of monolingual learners' dictionaries is that they contain detailed information on grammar and usage, a development which Rundell (1998: 337) views as "a welcome move away from the inappropriate model of the native-speaker's 'dictionary of record' towards a more 'utilitarian' lexicography, in which the needs of the user take precedence over all other factors". Although the LEAD was conceived in a different framework, it is fully compatible with Bergenholtz & Tarp's (2010: 29) *function theory*, which considers both general and specialized dictionaries as "utility tools conceived to satisfy specific types of human needs". For Tarp (2009: 47), lexicographic works need to be designed "with the genuine purpose of meeting the specific types of information needs which a specific type of users may have in a specific type of situation". In the case of the LEAD, the lexicographic resource is designed to meet the needs of intermediate to advanced non-native speakers of English writing academic texts either as students or as researchers.

This article is structured as follows: after presenting the general design of the LEAD, we describe the decisions we made to meet users' needs at the level of the macro-structure (Section 3), micro-structure (Section 4) and access structure (Section 5). In the concluding section (Section 6), we present some caveats and suggest a few avenues for future refinement of the tool.

2. General design of the *Louvain English for Academic Purposes Dictionary*

As can be seen in Figure 1, the LEAD is made up of the dictionary proper, which is the core of the resource, and two other components: a computer-assisted language learning (CALL) module and a corpus module. It therefore clearly qualifies as an “augmented reference tool”, a term coined by Nielssen (2010: 82) to refer to lexicographic resources which, besides the dictionary proper, also include separate sections with grammar, exercises, additional examples, etc. In this section, we present the general design of the LEAD by focusing on four of its main characteristics, i.e. the fact that it is *corpus-based*, *production-oriented*, *hybrid* and *customizable*.



1

Figure 1: General design of the LEAD

2.1. Corpus-based

Corpora play a key role in the LEAD. They were used to draw up the list of headwords (see Section 3) and to identify the information to be included, in particular the phraseological patterns of EAP words and the difficulties they generate for learners (see Section 4). Two types of corpus were used. First, the academic component of the *British National Corpus* (BNC) was utilized to describe the meaning and use of a selected list of academic words. This subcorpus is composed of approximately 10 million words of published academic prose representing a range of academic disciplines (humanities, social science, applied science, technology and engineering). Spoken data extracted from the spoken part of the BNC was also utilized sporadically to compare the use of some words or phrases in academic writing and speech. Second, a learner corpus was used to identify learners’ difficulties with EAP words. The learner data comes from the second edition of the *International Corpus of Learner English (ICLE)* (Granger *et al.*, 2009), which contains 3.5 million words of argumentative

essays written by high-intermediate to advanced learners of English from 16 mother-tongue backgrounds (Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Tswana and Turkish).

Corpora are also an integral part of the access structure of the LEAD. Users have direct access to discipline-specific corpora which can be searched for additional examples of the dictionary headwords and/or authentic contexts of use of words not included in the dictionary (see Section 5).

2.2. Production-oriented

EAP resources have generally catered more for users' receptive needs, in particular reading comprehension, than productive needs. Many studies have focused on the number of words learners need to know in order to achieve adequate comprehension of academic texts (cf. Nation 2001). In a writing perspective, however, it is not enough to know what the words mean; it is essential to be able to use them idiomatically and in this respect lexicographic resources are sadly lacking. And yet, as shown by Yihua & Qiping's (2010: 172-3) study of the use of specialized dictionaries in China, production-oriented information is the type of information that scores highest on students' wish lists. The best way to identify the productive needs of non-native EAP writers is to analyse their own writing. Learner-corpus-based studies have shown that one major – if not **the** major – source of difficulty was insufficient knowledge of the phraseological patterning of EAP words. This deficiency may result in miscolllocations of the types illustrated by examples (1-4) extracted from the *ICLE*, featuring the word *attention* used with an erroneous or atypical preposition and/or verb. Other attested difficulties stem from the routine-like nature of academic writing. Many EAP words occur in recurrent sequences, which Biber *et al.* (1999) call 'lexical bundles' (see Section 4.1.2). For example, typical lexical bundles with the verb *conclude* include *it can/may be concluded that*, *we may conclude that* and *it is/seems reasonable to conclude*. Failure to use these routine formulae results in clumsy formulations such as those illustrated in examples (5-9). It is this type of productive need that the LEAD seeks to cater for (for more examples, see Gilquin *et al.* 2007).

- (1) *businessmen pay more attention **on** money*
- (2) *Without focusing our attention **at** the university degrees of a particular faculty*
- (3) *Moreover, government needs to **put** more attentions on education*
- (4) *more people **get** their attention to health*
- (5) ***I would like to conclude that** it is good to have cyber cafes.*
- (6) ***I would conclude by saying** South Africa has best soccer players*
- (7) ***I want to conclude saying that** the Gulf war was impudently a war of interests*
- (8) *After discussing the above considerations, **it makes me to conclude that** recycling should be retained.*
- (9) *Evaluating the evidence **I come up to conclude that** football betting should be legalized.*

2.3. Hybrid

Developments in electronic lexicography are progressively breaking down the barriers between the different types of dictionaries and between dictionaries and other reference works. This phenomenon, referred to by Hartmann (2005) as 'hybridization', is of particular relevance for pedagogical lexicography: "a well-planned and executed hybridization may lead to a very effective lexicographic product with an expanded pedagogical function" (Gouws 2010: 56). The hybridization of the LEAD is twofold. First, it caters for both general academic needs (the dictionary proper) and discipline-specific needs (direct access to

discipline-specific corpora) on the basis that, as pointed out by Bowker (2010: 159), “many learners in specialized fields (...) may be non-native speakers who need help with both general and specialized language”. Second, the LEAD is both a reference and a learning tool. It perpetuates a tradition started by monolingual learners’ dictionaries to progressively encroach on “territory formerly occupied by mainstream English language teaching materials” (Rundell 1998: 337). Electronic lexicography has taken this trend one step further and generated new categories of hybrid tools, such as “dictionary-cum-CALL” and “CALL-cum-dictionary” (Abel 2010). The LEAD qualifies as dictionary-cum-CALL, as its learning component is subsidiary to its lexicographic function.

2.4. Customizable

Another key feature of the LEAD is that the content can be customized to users’ needs. This is made possible by its electronic format: “[t]he electronic dictionary is not restricted to specific user profiles but can be accommodated to the needs of different user categories. The user interface can be customized to fit the capabilities and needs of the individual users” (Svensén 2009: 446). It is important to bear in mind that the LEAD was conceived from the outset for a well-defined user base. Although aspects of the resource may be useful to novice native writers, its priority targeted users are non-native writers of high-intermediate to advanced proficiency levels. Three user groups can be distinguished: (1) undergraduate or graduate/postgraduate students who have to write academic texts (essays, reports, dissertations, etc.) in English in their chosen discipline (economics, sociology, linguistics, psychology, medicine, etc.); and (2) professionals in a particular subject field whose command of English is high but not quite sufficient to write fully idiomatic academic texts. These are “mature users who employ learners’ dictionaries to help them express their ideas in a foreign language that may have become their professional *lingua franca*. Such users are fluent and competent native speakers of another language, but they need contextualized dictionary information to produce adequate texts in a non-native language” (Bowker 2010: 158). A third group consists of teachers who can use the LEAD as a companion teaching resource in their EAP classes.

Although the LEAD’s user base is narrowly defined, it is still possible to bring the tool closer to users’ needs thanks to flexible data presentation (Trap-Jensen 2013: 44). Two types of customization of particular relevance for EAP are implemented in the LEAD. First, the content is partially adapted to a user’s native language, thereby making it possible to pinpoint difficulties that are specific to a particular mother-tongue background (for more details, see Section 4.2). Second, the examples are automatically extracted from corpora representing the user’s selected discipline (see Section 4.1.3). While this represents a step in the direction of greater user fit, the degree of customization of the LEAD is admittedly still quite modest: only part of the content is customizable and the tool qualifies as “adaptable” rather than “adaptive” (Gamper & Knapp 2002), i.e. it requires manual action on the part of the user (selection of mother-tongue background and discipline) rather than automatically adjusting to user input.

3. Macrostructure

The macrostructure of the LEAD is based on the *Academic Keyword List* (AKL), a list of 930 potential academic words, i.e. “words that are reasonably frequent in a wide range of academic texts but relatively uncommon in other kinds of texts and which, as such, *might* be used to refer to those activities that characterize academic work, organize scientific discourse and build the rhetoric of academic texts, and so be granted the status of academic vocabulary”

(Paquot, 2010: 29). The method used to extract potential academic words primarily relies on keyword analysis, a fully data-driven technique for identifying distinctive words or keywords that consists in comparing the frequencies of words in the corpus under study with their frequencies in a reference corpus. For the purpose of compiling the AKL, corpora of professional and student academic writing were compared with a large corpus of fiction on the grounds that academic words would be particularly under-represented in literary texts. Keywords were extracted with the Keyness module of *WordSmith Tools 4* (Scott, 2004). This method, however, cannot distinguish between ‘global’ keywords dispersed more or less evenly through the corpus and ‘local’ keywords that appear repeatedly in some parts of the corpus only (Katz, 1996). Consequently, the criteria of range and evenness of distribution were subsequently used to filter the resulting list of potential academic words and narrow it down to a list of the more global keywords. Range (i.e. the number of texts in which a word appears) was used to determine whether a word appeared to be a potential academic keyword because it occurred in most academic disciplines or because of a very high usage in a limited subset of texts. Unlike range, evenness of distribution is a statistical coefficient (here Juilland’s D) that takes into account “not only the presence or absence of a word in each subsection of the corpus, but the exact number of times it appears” (Oakes and Farrow, 2007: 91). This criterion was used to distinguish between two types of words that appear in a wide range of texts and disciplines, i.e. words such as *example* which we intuitively regard as an academic word, and words such as *law*, the meaning of which is more discipline- or topic-dependent (e.g. *canon law* in theology, *criminal law* in law, *the law of gravity* in physics) (for more details about the AKL compilation, see Paquot, 2010: 29-63).

The AKL differs substantially from Coxhead’s (2000) *Academic Word List* (AWL), i.e. the most widely used academic word list in language teaching, testing and materials development today. The AWL was developed to meet the specific receptive vocabulary needs of students in higher academic settings: taken together, the 2,000 most frequent words of English as described in West’s (1953) *General Service List* (GSL) and AWL words plus technical items should approach the critical 95% coverage threshold necessary for reasonable reading comprehension (Nation 2001: 197). Unlike the AWL, the main objective of the AKL was to fulfil the productive needs of students; it is likewise not based on an *a priori* division of vocabulary into mutually exclusive lists of core, technical and academic words. The procedure described above resulted in a list of 930 potential academic words, some 60% of which are GSL words (e.g. *aim, argue, because, compare, explain, namely, result*); only 40% of AKL words also occur in Coxhead’s AWL.

Nouns make up almost 40% of all potential academic words included in the AKL, a large proportion of which are illocutionary nouns (*question, argument*), language-activity nouns (*contrast, definition*) and mental process nouns (*analysis, hypothesis*). Verbs account for a quarter of the list and mainly include activity verbs (*use, show*), reporting verbs (*discuss, explain*), mental verbs (*examine, interpret*), linking verbs (*appear, become*) and logico-semantic relationship verbs (*cause, contrast, follow*). Although usually not the main focus of academic textbooks and teaching materials, adjectives represent about 20% of the AKL; they often express value judgments (*appropriate, clear, useful*), possibility (*potential, likely*), and logico-semantic relationships (*alternative, different, following, parallel, similar*). Adverbs (approximately 10%) consist essentially of linking (*consequently, hence, thus*) and evaluative (*correctly, effectively, inevitably*) adverbs. The AKL also includes prepositions and conjunctions, many of which are complex forms (e.g. *such as, according to, because of, given that*).

The *Academic Keyword List* is not a final product. It requires “pedagogic mediation” (Widdowson, 2003): each AKL word was subject to a careful corpus-based analysis to confirm its status as an academic word. Since the beginning of the project, we have also extended the coverage of the dictionary, most notably on the basis of users’ queries and no-match lists from the log files (cf. Trap-Jensen 2013: 43): the list of lexical entries now numbers approximately 1,200 and is still growing.

4. Microstructure

In view of the overall aim of the LEAD, the dictionary microstructure mainly includes two types of information: a detailed description of the phraseological patterning of EAP words (Section 4.1) and other types of usage information directly linked with learners’ attested difficulties (Section 4.2). These elements are all stored in different tables in a MySQL database and the data presentation structure is programmed in PHP and JavaScript.

4.1. The phraseological patterning of EAP words

The preferred company of EAP words in terms of both collocations and lexical bundles is given pride of place in the microstructure of the LEAD. Like Prinsloo (2009:182), we strongly believe that “the contribution of the corpus of the future and of the future of the corpus depends on the extent to which comprehensive behavioural patterns of words can be compiled and presented to lexicographers in a condensed but user-friendly, machine-readable format”. This section describes the extraction and presentation layout of collocations (Section 4.1.1) and lexical bundles (Section 4.1.2) and highlights the key role played by examples (Section 4.1.3).

4.1.1 Collocations

Collocations were extracted automatically from the BNC academic sub-corpus using the Word Sketch tool of the *Sketch Engine* (Kilgarriff *et al.*, 2004; Atkins & Rundell, 2008: 107-11). This tool provides a list of collocates for a wide range of grammatical relations in which a given word participates. For the noun *issue*, for example, Word Sketch includes relations such as ‘modifier’ (e.g. *important, major, political, central issue*), ‘object of’ (*raise, address, discuss, consider + issue*), ‘subject of’ (*issue + arise, relate, surround*), ‘adj subject of’ (i.e. predicative use of adjectives, e.g. *issue + important/concerned/complicated/complex*), ‘and/or’ (*issue + and/or + problem/question/theme*) and a variety of preposition-based relations (e.g. *issue + of + policy/principle/importance/concern; number/range/discussion/complexity + of + issue*), some of which seem to cross part-of-speech (POS) categories for collocate extraction (e.g. the ‘pp_obj_to-p’ relation features collocates such as *attention/relevant/relate + to + issue* while the ‘pp_obj_on-p’ relation includes patterns such as *focus/view/concentrate/divide/debate + on + issue*).

We restricted the number of grammatical relations to what we believe are the most salient and useful collocation categories for academic writing purposes: adjective + noun, verb + noun, noun + verb, adverb + adjective, and adverb + verb. The preferred prepositions to be used with a given academic word are currently recorded in the form of lexical bundles (see 4.1.2).

Figure 2 is a typical collocation box for a noun in the LEAD. It features adjective + noun collocations, verb + direct object collocations and subject + verb collocations of the noun

assumption. Collocations are sorted by order of decreasing frequency in the BNC academic sub-corpus.

Collocations
Adj + <i>assumption</i> <i>basic, certain, underlying, implicit, fundamental, common, general, reasonable, theoretical, different, false, key, important, hidden, shared</i>
V + <i>assumption</i> <i>make, challenge, question, share, accept, hold, reject, use, justify, examine, reflect, test, support</i>
<i>assumption</i> + V <i>underlie, seem, appear, lead, govern, lie</i>

Figure 2: ‘Collocations’ menu for the noun *assumption* in LEAD

4.1.2. Lexical bundles

Learners’ difficulties are not restricted to collocations. EAP also has its typical routines, i.e. preferential word sequences such as *the point is that*, *as described above* and *stands in marked contrast to*. These sequences are all the more difficult to notice – and therefore subsequently use – as they tend to be semantically compositional and therefore not cognitively salient (for the difference between cognitive and social salience, see Hanks 2013: 343). A recent experiment conducted by Jalali (2014) has highlighted learners’ difficulty in selecting the typical word sequences. Our own research based on learner corpus data confirms learners’ lack of awareness of these routine formulae. For example, learners are often found to use the verb *argue* in active sequences like *people argue that*, *some/many people argue that* or *one can argue that* instead of the more typical passive patterns (*it has been argued that*, *it can/could/might be argued that*, *as argued by*). These lexical bundles were extracted automatically from the BNC academic sub-corpus using the ‘clusters’ option of the ‘concord’ tool in WordSmith Tools 5 (Scott, 2008). This method generated long lists of two- to five-word sequences ordered in decreasing frequency, which we then systematically checked manually. Grammatically incomplete combinations were deleted (e.g. *issue of the*, *this issue has been*) as well as embedded sequences (e.g. *significant difference in* was deleted, as all occurrences were embedded in the larger sequence *a significant difference in*).

By default, five lexical bundles are displayed in the right-hand menu. As illustrated in Figure 3, by clicking on the three dots at the bottom of the ‘Lexical bundles’ frame, the user has access to a longer list of lexical bundles in a pop-up window. Like collocations, all lexical bundles are hyperlinked to a maximum of five discipline-specific examples (see Section 4.1.3).

The screenshot shows the 'The Louvain EAP dictionary' interface. At the top, there is a search bar with the text 'Search Selected discipline: Linguistics; selected mother tongue: French' and a 'Help' link. Below the search bar are navigation tabs: 'Home', 'Search by function', 'Search by translation', 'Corpus search', and 'Exercises'. The main content area displays the dictionary entry for 'issue (n.)', which includes a definition: 'an important subject that people discuss and have opposing views about (see also [question](#)):' and an example sentence: 'This chapter also addresses **the issue of** consent.' Below the definition is a 'Lexical Bundles' pop-up window with a list of phrases such as 'there is the issue of', 'the issue here is', 'what is at issue is', 'a wide range of issues relevant to the issue of', 'this is an important issue', 'that is not the issue', 'the point at issue is', 'many of the issues raised', 'one of the key issues', 'perhaps the most important issue', 'we shall return to this issue later', and 'to consider the issue of'. To the right of the main entry, there are sections for 'Collocations' (Adj + issue, V + issue) and 'Lexical bundles'.

Figure 3: The ‘lexical bundles’ pop-up window

Presenting these sequences explicitly to learners might be criticized on the grounds that it may result in overuse and a generally stale, cliché-ridden style. However, as aptly pointed out by Henry (2000: 124), a distinction needs to be made between native and non-native speakers on this issue: while native speakers may need to avoid clichés, learners “need to acquire a larger phrasal lexicon, in other words, more clichés”. Based on our own pedagogical experience, we concur with Jalali’s (2014: 8) observation that “students, unlike published writers, need more exposure and practice in the use of these building blocks of discourse”. Including a list of frequent bundles in a pop-up window appeared to be a convenient way of raising learners’ awareness of an important aspect of EAP formulaicity that is still largely disregarded in EAP resources.

4.1.3. Examples

Examples are a key component of any dictionary entry. The controversy that opposed proponents of invented vs. authentic examples (cf. Laufer 1992) has lost much of its relevance, as today all dictionary examples are, to a greater or lesser extent, based on corpus data (Rundell 1998: 334). The main issue is no longer whether to use corpora but rather which corpora should be used. Ideally, the variety of language represented in the corpus should match the user’s writing needs as closely as possible. In the case of specialized dictionaries, Nielsen (2014: 193) advocates the inclusion of “domain-specific example sentences illustrating how source language conventions and style can be transposed to a foreign language”. In a cross-disciplinary academic dictionary like the LEAD, we could have made use of a large domain-undifferentiated EAP corpus. We opted instead for a series of domain-differentiated corpora and implemented a system that automatically customized the examples to the user’s selected discipline. It seemed reasonable to us to assume that it would be more motivating and useful for users to be exposed to examples from their own field. When you write a text on chemistry, an example from medicine or psychology may seem very distant from your own concerns and may also pose some challenges in terms of intelligibility. A recent study by Chang (2014) appears to validate our decision. She compared Korean engineering students’ experiences of consulting general and specialized corpora for academic English writing and found that they had a clear preference for the specialized corpus “for its direct relevance to their academic fields” (p. 254).

The Louvain EAP dictionary interface shows the entry for *factor* (n.). The definition is: "a fact or situation, often one of several, which influences the result of something:". Two example sentences are provided: "Scientists think that global warming may be a **key factor** in the increase of disease in frogs." and "A **second major factor** was the link between the army and the rural community." A box in the main content area displays five examples of the collocation *identify + factor* from a psychology corpus: "To **identify** potential **factors** associated with concealment and disclosure, analysis of variance and eta-square (for categorical variables) were used to examine mean differences, and the Pearson correlation coefficient was used to examine correlates (for continuous variables) at the bivariate level, followed by simultaneous linear regression of all factors to identify the unique association of each factor to disclosure and concealment. (PSY)", "They also **identified factors** such as age and its typical concomitants of family, career, and material concerns that shift individual values from less ideological to more pragmatic over time. (PSY)", "The most prevalent taxonomy of individual differences **identifies** five broad personality **factors** (Digman, 1990). (PSY)", "The convergent findings of altered stress response and altered pain regulatory mechanisms in smokers are novel and of specific importance in terms of **identifying factors** contributing to maintenance of nicotine dependence. (PSY)", and "Researchers have **identified** several **factors** that influence the relationship between turnover rates and organizational performance (Hausknecht & Trevor, 2011; Shaw, 2011). (PSY)". The right-hand column lists collocations: "Adj + factor: other, important, major, significant, main, key, different, relevant, associated, contributing, various, crucial, additional, external, causal, only, contributory, underlying", "V + factor: identify, consider, involve, relate, find, know, examine, include, become, mention, show, suggest, control, establish, discuss, list, describe, measure", and "factor + V: influence, affect, contribute, determine, cause, lead, explain, play, relate, account, operate, increase, shape, tend, bind, constrain, need, inhibit, involve, vary, underlie, act, regulate". The Lexical bundles section shows "factor in", "factor of", and "factor for" with the example "is an important factor in one of the factors that".

Figure 4: The entry for *factor* in the LEAD

The dictionary interface is programmed to provide discipline-specific examples of collocations and lexical bundles. As illustrated in Figure 4, by clicking on any collocate of the noun *factor* in the right-hand column, the user is shown five examples of the selected word combination (here *identify + factor*) in a box that opens in the main frame. Examples are automatically selected from a corpus that represents the user's selected discipline (here psychology). If the user had selected medicine as his or her main discipline, examples such as the following would have been displayed instead:

10. The study thus **identified** several **risk factors** apart from treatment, which included preexisting ILD, which were not treatment specific, and which were partly similar to risk factors for idiopathic or rheumatic pulmonary lung fibrosis. (MED)
11. The study can **identify factors** that might be associated with reporting in the target population, but access to health care is likely to be an independent risk factor for underreporting in the general population. (MED)

Similarly, the lexical bundle *is an important factor in* is illustrated by different corpus examples according to the user's selected discipline. Examples 12 to 14 illustrate the type of authentic examples automatically retrieved for education, linguistics and business.

12. These results provide new evidence that the prevalence of peer victimization in high school **is an important factor in** high school academic performance. (EDU)
13. While we wait for future studies to address the issues raised in this section, the findings of this study provide additional support for previous research suggesting that topic familiarity **is an important factor in** L2 reading comprehension. (LING)
14. Accumulated exploration cost, as introduced earlier as *ct* for time period *t*, **is an important factor in** deciding whether or not an entrepreneur should continue exploration. (BUS)

As the discipline-specific corpora available in the LEAD are of a relatively limited size, if five examples of a collocation or lexical bundle are not found in the user's selected discipline-specific corpus, more examples are extracted from the other discipline-specific corpora and displayed in order of 'domain proximity'.¹ Thus, if psychology is selected as the preferred discipline, the system will query the psychology corpus to retrieve five discipline-specific examples; if it does not identify five examples, the system will then query the remaining discipline-specific corpora following a set path that starts with the most closely related disciplines (1. Education, 2. Sociology, 3. Anthropology). By contrast, if law is selected, the system will proceed from the law corpus to corpora representing political science, economics and business. Domain proximity was established on the basis of our understanding of the various disciplines and their interactions. We still need to validate the different set paths with domain experts.

Examples were extracted fully automatically from the discipline-specific corpora. At first fully unsupervised, the method was progressively fine-tuned in order to weed out a number of unsuitable examples. The first problem that needed to be fixed was the mismatch between the automatically extracted examples and the syntactic relations they were supposed to illustrate. This usually resulted from the fact that the corpora used to extract collocations were POS-tagged but were not parsed. To remedy this problem, we used regular expressions to match instances of collocations in a given syntactic relation. To extract verb-noun collocations in a direct object relationship, for example, we restrict the search window to four words to the right of the verb and use POS filters to block collocates separated by a pronoun, a conjunction, a punctuation mark, etc. Another category that needed to be filtered out concerned examples that were felt to be too difficult to process. Inspired by Kilgarriff and colleagues' GDEX tool, we are currently working on the implementation of selection rules that prioritize shorter examples with a limited number of punctuation marks, numbers, special symbols and capitals (cf. Kilgarriff *et al.*, 2008).

a. Other types of information

The LEAD is grounded in recent research that has highlighted the benefit of form-focused instruction for vocabulary learning (Laufer and Girsai, 2008). The dictionary explicitly draws users' attention to a range of problems that non-native speakers experience when writing academic texts. Careful investigation of authentic learner data from the *International Corpus of Learner English* (Granger *et al.*, 2009) has helped uncover a range of difficulties involving semantic misuse, lack of register awareness, atypical syntactic patterning and problems of frequency (what has been referred to as 'overuse' and 'underuse' in learner corpus research), all of which are addressed in the LEAD. For example, the LEAD contains an entry for the adverb *of course* but there is a note that warns learners against its use in academic writing. As illustrated in Figure 5, this warning is reinforced by a chart showing that the frequency of the adverb in learner writing is closer to its frequency in speech than in academic prose.

¹ In previous versions of the LEAD, further examples were extracted from the academic subcomponent of the British National Corpus (cf. Paquot, 2012).

of course (*adv.*)

used to emphasize that something is very clear or already known

*The upheavals of the Revolutionary era in France allowed some remarkable individuals to rise from lives of obscurity to positions of fame. Napoleon, **of course**, is the greatest example of this phenomenon.*

Be careful!

Learners often use the adverb **of course** to strengthen their claims but the adverb is more frequent in speech than in writing.

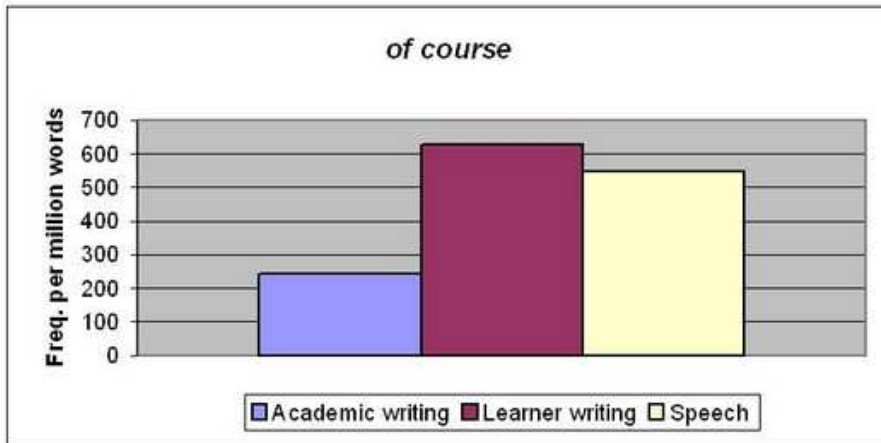


Figure 5: The entry for *of course* in LEAD

Common learner errors are addressed in warning boxes which typically provide a description of an attested learner error, followed by an indication of the form that should have been used instead. The error is illustrated in an authentic learner sentence and corrected. The lexical entry for the conjunction *as*, for example, includes a warning against learners' tendency to use the impersonal pronoun *it* in subject position after *as* (Figure 6).

Warning

Do not use the impersonal pronoun **it** in subject position after **as**:

- *Children are not playing marbles any more, **as it was** the case about thirty years ago.*
- *Children are not playing marbles any more, **as was** the case about thirty years ago.*

Figure 6: Warning note under the entry for the conjunction *as* in the LEAD

Errors and difficulties found in the writings of a wide range of learner populations are dealt with in generic notes that are displayed irrespective of the L1 background selected by the user. Other notes, however, target the problems that a specific L1 learner population (for the moment, French learners only) typically encounters. One of the purposes of L1-background customization in the LEAD is to provide contrastive feedback and draw learners' attention to major differences between English academic words and their potential translation equivalent forms in the learners' L1. Thus, if French is selected as the user language and the user searches for the French verb *prétendre* in the semi-bilingual access mode (see Section 5), the LEAD offers its most frequent translation (i.e. *claim*) as well as a note that warns French users not to translate *prétendre* as *pretend*.

For more details about learners' attested difficulties and their treatment in the LEAD, see Gilquin *et al.* (2007) and Paquot (2012).

15. Access structure

Electronic lexicographic tools explode the traditional structures of dictionaries. As put by Svensén (2009: 441), “[s]everal traditional structures have become less important or even quite irrelevant, and the structure predominating in an electronic dictionary is the access structure”. Like that of a print dictionary, the access structure of an electronic dictionary consists of the organization of the indicators directing users to the information they are looking for in the dictionary. However, electronic dictionaries have much more powerful types of search and linking facilities than print dictionaries.

Search and linking facilities are entry points to the lexicographic data that need to be modelled with a view to satisfying specific users’ requirements: “When planning the access structure of a specialized monolingual dictionary for learners the lexicographer needs to determine the typical consultation route of the intended target user when trying to reach an answer for a typical question that directs him/her to the specific dictionary” (Gouws, 2010: 61) (see also Müller-Spitzer, 2013). In the LEAD, search facilities meet three types of users’ needs. First, the dictionary can be used as a standard semasiological learner dictionary and provide help to users who think of a target word or phrase but are unsure about its meaning or use (What is the exact meaning of *nevertheless*? What is its typical position in a sentence? Which verbs and adjectives collocate with the noun *implication*?). The LEAD is also a “semi-bilingual dictionary” (Laufer and Levitzky-Aviad 2006) or a “bilingualized L2 dictionary” (Svensén, 2009: 383), i.e. users who select a particular mother-tongue background (currently Dutch and French) can perform a bilingual search to access English lexical entries via an index cross-referencing from the L1 equivalents to the English lemmas. The third access mode aims to provide help to users who may know only one or two ways of expressing a particular function (e.g. contrasting two arguments or expressing a cause-effect relationship) and would like to vary the linguistic devices used to perform this function (cf. also Pecman 2008; Kübler and Pecman, 2012; L’Homme *et al.*, 2012). While in other dictionaries that provide this type of access the onomasiological search system is usually based on semantic relationships, in the LEAD it relies on a pull-down menu of 18 rhetorical or organizational functions that we identified as being particularly prominent in academic discourse (e.g. compare and contrast; define terms; express cause and effect; express personal opinion; introduce a concession; introduce a topic; list and sequence; quote and report; refer to tables, figures and graphs; reformulate). 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. Figure 7, for example, provides a list of linguistic items that may be used to express possibility and certainty. One of the main advantages of this access mode is that it suggests alternatives and thereby helps users expand their academic repertoire. Figure 7 shows that *perhaps* and *maybe* are not the only adverbs that can be used to express possibility; nouns (e.g. *chance*, *likelihood*, *possibility*), verbs (e.g. *appear*, *seem*, *suggest*) and adjectives (e.g. *apparent*, *likely*, *possible*) are alternative options.

CECL The Louvain EAP dictionary

Express possibility and certainty Search [Help](#)

[Home](#) [Word search](#) [Search by translation](#) [Corpus search](#) [Exercises](#)

Express possibility and certainty

Nouns	Verbs	Adjectives	Adverbs	Modals
assumption belief certainty chance likelihood possibility probability tendency	appear prove seem suggest tend	apparent certain clear doubtful essential evident impossible likely necessary obvious possible probable reasonable unlikely	absolutely apparently certainly clearly completely consistently essentially exclusively inevitably maybe naturally necessarily obviously of course perfectly perhaps possibly presumably probably really undoubtedly	can could may might must should

Figure 7: Onomasiological search system in the LEAD

Linking, another important access facility in the LEAD, takes the form of various dictionary-internal and dictionary-external hyperlinks. There are two main types of dictionary-internal links, which can be used in the classroom and/or for autonomous learning. First, the semasiological and onomasiological access structures “intertwine and support each other” (Jónsson, 2009: 259) in the LEAD and users can navigate from a lexical entry to a list of other linguistic devices that are typically used to perform the same rhetorical or organizational function by clicking on the Function link in the left-hand menu. The noun *belief*, for example, is hyperlinked to two functions, i.e. ‘Expressing possibility and certainty’ and ‘Quoting and reporting’. Thus, after reading that a *belief* is a ‘strong feeling that something is true or exists’, a user may click on ‘Express possibility and certainty’ to find other nouns that may be used to express a lower degree of certainty (*assumption, chance, likelihood, possibility*), as well as verbs (*appear, seem, suggest*), adjectives (*likely, possible, probable*) and adverbs (*apparently, presumably, probably*) that may also be used to perform this specific function; the user may also decide to click on ‘Quoting and reporting’ to find alternative ways to report an author’s ideas (e.g. *argument, assumption, claim, contention, hypothesis, remark, statement, suggestion*).

The second type of dictionary-internal link serves to connect lexical entries and related CALL exercises targeting learners’ attested difficulties. Exercises such as fill-in-the-blanks, multiple-choice exercises, word building and collocation exercises were created using the Hot Potatoes software (<<http://hotpot.uvic.ca/index.php>>) and are fully integrated within the dictionary online environment. Exercises may target specific lexical items that have been found to be particularly error-prone and for which an error note is also available. Thus, the LEAD not only describes the common confusion between *on the other hand* and *on the contrary* or between *i.e.* and *e.g.* in error notes, but also offers exercises that target these specific problems. In multiple-choice exercises, the distractors are not chosen at random but include attested errors and items that are typically confused by learners (see also Usami 2013). As such, multiple-choice exercises play an active part in the learning process. Other exercises

cover a whole range of lexical items typically used to express a specific function. These exercises are not hyperlinked to specific lexical entries but are available when a user selects a specific function in the pull-down menu of rhetorical and organizational functions. While most of the 60 exercises currently available in the LEAD are potentially useful to all users, we also included exercises that deal with French learners' specific difficulties in the form of 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 given French word (e.g. the different translations of French *comme* into English: *as*, *like* and *such as*).

A corpus tool is also available via two types of dictionary-external links that fulfil two different types of users' needs (Paquot, 2012). First, a dictionary user may wish to have access to more examples of academic words and phrases in context, thus expanding on dictionary information through corpus consultation. To respond to this particular need, the hyperlink under 'More corpus examples from ...' at the bottom of the right-hand menu in a lexical entry (e.g. the verb *differentiate*) is programmed so that by clicking on the selected discipline (e.g. linguistics), the user can directly access authentic examples of the verb *differentiate* in a corpus of linguistic articles, as illustrated in Figure 8. With this first type of dictionary-external link, we are trying to implement Cobb's (2003) view of "the ideal electronic resource for language learning" that would consist in "a blend of dictionary and concordance" and allow "searches of the dictionary publisher's corpus from individual entry pages". Second, a dictionary user may want to check in a discipline-specific corpus how to use a word that is not covered in the LEAD. These may be general English words or discipline-specific terms. To meet this type of need, we have made the corpus query tool available as an independent component in the form of a button in the navigation bar (cf. Figure 8). We opted for CQPweb, an open source web-based corpus handling tool (Hardie, 2009) that handles annotated corpora (corpora available in the LEAD are lemmatized and part-of-speech tagged) and offers powerful search syntax. Users can not only search for word forms or lemmas (e.g. all the word forms of the verb *overrule* in the law corpus) but they can also make use of regular expressions to extract word sequences (e.g. searching in the law corpus for all word sequences that match the pattern *the court* + past tense of lexical verb retrieves the following sequences: *the court held, found, concluded, noted, reasoned, rejected, ruled, stated, decided, reiterated, acknowledged, observed, considered*). The tool also offers facilities such as sorting options and automatic collocation extraction.

In the LEAD, users can decide to look up words in the academic component of the *British National Corpus* or in discipline-specific corpora of one million words each (currently anthropology, business, economics, education, law, literature, linguistics, medicine, political science, psychology and sociology). Research on data-driven learning in the L2 academic writing class has shown that corpora are useful reference tools for problem solving in writing (Yoon, 2011). Students have also been reported as saying that they appreciate the value of scanning multiple authentic examples derived from a user-centred corpus that serves as a writing model (e.g. Lee & Swales, 2006; Rodgers *et al.* 2011). This makes it all the more essential that "thorough and detailed needs analysis should precede the compilation of a specialized corpus" (Chang 2014: 255). We therefore compiled discipline-specific corpora in close collaboration with experts in the different disciplines, who recommended a shortlist of top-tier international journals that could be used as high-standard writing models by students and researchers. When a discipline-specific corpus is selected, users can also restrict their search via the 'Restrict query' option to highly specific components of the corpus representing sub-disciplines such as respiratory and critical care medicine or applied linguistics. Users can also select journal titles; this option is particularly useful for researchers who plan to submit a

manuscript to a specific journal and wish to familiarize themselves with its language and style.

of six months . The study focuses on 15 phonological features that	<u>differentiate</u>	British English (BrE) and AuE . Each one involves pronunciations
was post-vocalic /r/ , as in car , a feature that categorically	<u>differentiates</u>	North American English (NAmE) , which is rhotic , and
the Canadian children . The important observation is how the children are	<u>differentiated</u>	. The 9-year-old uses 100 British (t) . However ,
Kanji processing . However , their basic understanding is hardly sufficient for	<u>differentiating</u>	valid from invalid radical information and incorporating valid information se
) in the language , together with a single threshold probability to	<u>differentiate</u>	strings as grammatical (above the threshold) or ungrammatical (below
1b) a possible positive reading : <gap> (1) Biq further	<u>differentiates</u>	the metalinguistic meaning of the negation marker mei from its descriptive ,
displaying how he uses Maria 's accent as a basis of	<u>differentiating</u>	himself from her : "Cause I do n't say it with any
with a phonological form . A criterion must specify a way of	<u>differentiating</u>	such cases from systematic uses . What is required is not simply
both omissions and overextensions . However , the TLU score does not	<u>differentiate</u>	between omission and overuse , which mean quite different things in acquisit
of the EC would be strengthened if it were shown to consistently	<u>differentiate</u>	learners before and after the treatment . This leads to so-called 'predictive
that in the case at hand , some linguistic factor or factors	<u>differentiating</u>	the speech of the two speakers is responsible for the lack of
respects . First , the Estonian case system is larger and more	<u>differentiated</u>	, with 14-odd cases divided between a predominantly fusional grammatical st
primary tense . Theme and logical interdependency . We then began to	<u>differentiate</u>	the occurrences selected for research . The projecting clauses , through their
the notion of language use may not have been made explicit or	<u>differentiated</u>	enough , and this suggestion is supported by judges ' comments .
(i.e. , to distinguish between different subcategories (e.g. , to	<u>differentiate</u>	a coffee cup from a tea cup ; see Clark , Gelman
constructions in a similar manner , suggesting that the children do not	<u>differentiate</u>	between the two structures (Nicoladis , 2001) . The most
.22 , p= .236) . Therefore , these analyses can not	<u>differentiate</u>	whether it was the English or French family size that had an
) sees Middlesbrough youth participating in a national trend and serves to	<u>differentiate</u>	the accent from that of Tyneside . As the Yorkshire orientation has
So she , like Jefferson , is saying that not all speakers	<u>differentiate</u>	between tokens (though qualified by " at least , " suggesting
BIOGRAPHIC NARRATIVES IN APPLIED LINGUISTICS Narrative study	<u>differentiates</u>	between three interconnected types of information one might gather from life
But there is currently no evidence that they use these dimensions in	<u>differentiating</u>	among words within the same semantic domain when choosing vocabulary .

Figure 8: Authentic examples of the verb *differentiate* in linguistics

16. Conclusion and perspectives

Although, as aptly pointed out by Tarp (2009: 61), “[n]obody knows how dictionaries will develop in the future”, we fully subscribe to his prediction that “one direction will probably be the ‘individualization’ of the lexicographic product, adapting it to the concrete needs of a concrete user in a concrete situation and providing much quicker and easier access to the relevant data”. The LEAD cannot claim to have achieved this ambitious aim, but it is a step in the right direction: it does not go so far as to cater for the needs of an individual user but focuses pointedly on those of a well-defined group of users faced with similar challenges.

The LEAD is also in line with Nielsen’s (2013, 370) prediction that “dictionaries of the future will, to an increasing extent, be regarded as ‘digital assistants’”. It is one of a new generation of tools being developed to meet users’ language production needs. Like its forerunner for French, the *Base Lexicale du Français* (Verlinde & Binon 2009), the LEAD functions as both a writing aid and a learning tool and includes a range of external components (corpus, exercises, etc.) in addition to the dictionary proper.

Although the LEAD is usable and, indeed, is currently being utilized by a large number of students and researchers on the Louvain intranet, there is still plenty of scope for further development, in particularly with regard to the number of EAP words and disciplines (currently mainly humanities). The customizability of the dictionary could be improved by adding more mother-tongue backgrounds and its general usefulness would benefit from a larger number of usage and error notes and exercises (in particular, involving collocations). We are also keen to make the corpus component of the tool more user-friendly by integrating a more intuitive concordancer. The whole issue of the use of concordancers by learners is controversial. While Gabel (2001: 269) argues that “concordancers are superior to traditional grammar books, dictionaries and coursebooks, because they allow easy access to huge amounts of ‘real’ language in use (...)”, Kilgarriff (2009) is more sceptical and argues that

“[m]ost learners do not want to be corpus linguists, and concordances are unfamiliar and difficult objects”. Although CQPWeb is an excellent and highly powerful tool, like most concordancers it was conceived with language specialists in mind. As the LEAD targets researchers in all disciplines, it should be designed in such a way that even non-linguists could easily retrieve the requested information.

More importantly, there is a crucial need for validation: we need to know how the dictionary is used and what users think of it. On the issue of customization, Trap-Jensen (2013: 45) rightly reminds us that there may be a huge gap between apparently ideal lexicographic decisions and actual dictionary use: “In a way, it is the lexicographer’s dream, but it has turned out to have one serious disadvantage: so far, studies have not been able to confirm that users take advantage of the possibilities offered to them”. Validation of the LEAD will need to distinguish between two different types of use: autonomous use where users are left to their own devices and classroom use when they are guided by their EAP teacher. Part of this validation work is already underway at the University of Louvain and the results are encouraging. Log files reveal that the dictionary is regularly used by the Louvain community, with 2,481 individual queries between 1 January and 24 March 2015. The LEAD is mostly consulted as a semasiological dictionary (about 70% of queries), but the dictionary is also often searched by function (about 25%). In addition to the implicit feedback gathered through log files (cf. De Schryver, 2013), we have also solicited users’ explicit feedback in the form of an online questionnaire that students (and colleagues) in the modern languages department were invited to complete. We still need to collect more questionnaires and systematically analyse the answers but it already gives us great satisfaction to see that a large majority of students (about 80%) answered the following question in the affirmative:

‘Je pense que j’utiliserai cet outil à l’avenir lorsque je devrai rédiger un texte académique en anglais’
(I think I’ll use this tool again when I need to write an academic text in English).

Students appreciate the many aspects of the LEAD microstructure (definitions, collocations, lexical bundles, usage notes and error notes) and also find the hyperlinked discipline-specific examples useful. Not very surprisingly (see discussion above on CQPweb’s user-friendliness), direct access to a concordancer is not highly rated at this stage by our students. Note, however, that, unlike students who often checked the ‘no opinion’ box, the eleven PhD students and researchers who completed the questionnaire agreed with the statement that access to a concordancer is useful. Interestingly, CQPweb log files reveal that at least some of our colleagues use the concordancer to search for word combinations such as *hardly occurred*, *little funding*, *careful revision*, *beyond the scope*, *provide evidence to support* and *this model can be used*. These preliminary results show that explicit feedback via a questionnaire and implicit feedback from log files are instrumental in identifying the LEAD’s most popular features and gaining a better understanding of how the different aspects of its microstructure and access structure are used by different groups of users. We therefore consider both types of feedback crucial for future development of the tool.

A longer-term but no less important development is the integration of the EAP tool into a wider package that would also include terminological resources so that users’ general and discipline-specific academic needs can be met in the same writing environment. Gouws (2010: 66) argues in favour of this type of integration: “Both types of learners, i.e. the learner of the subject field and of a language, can receive additional support by not only dealing with the dictionary in isolation but as part of a more comprehensive study package”. Having to use

two different tools, one for general needs and another for specific needs, is cumbersome for learners and can prove quite difficult, because users cannot be expected to draw a clear line between the two types of needs. As rightly pointed out by Bowker (2010: 166-7), one of the main strengths of hybrid dictionaries is that they allow for “one-stop shopping”: “This type of hybrid dictionary, which combines specialized terms with relevant and frequently used general language words, would be of great value to many types of LSP learners because it could provide them with a sort of ‘one-stop shopping’”.

A final and particularly challenging development aims to bring the tool closer to users’ writing environment. In its current state, the LEAD is a tool that users are expected to consult when they have a particular need. The problem is that there are many difficulties of which learners are unaware. For example, if a French-speaking user is unaware of the fact that the meaning of *on the contrary* only partially overlaps with that of *au contraire*, he or she is not likely to consult the dictionary to find out. A recent study by Chang (2014: 255) focused on corpus use confirms the reality of this danger: “The findings also point to the need for a certain degree of language proficiency for effective corpus consultation. During independent corpus use, some participants did not even attempt to look up their errors in a corpus because they believed that they were correct.” What is needed is automatic highlighting in the user’s text of the different types of information provided in the different structures of the dictionary. During or after writing a first draft, users should have the possibility of highlighting all the academic words used in the text, checking their definition, identifying their collocations and/or lexical bundles or preferred positioning, highlighting error-prone words, calling up other words and phrases that express similar functions, etc. In this kind of environment, the front-end is the user’s text; the dictionary is the back-end. The links go from the user’s text to the dictionary, rather than the other way round. Inspiration for the design of this interface could be drawn from De Wachter *et al.*’s (2014: 20) writing aid tool for Dutch: “*Writing Aid Dutch* guides students through their writing process by making them aware of the most frequent writing problems in their texts situated on the level of text structure, style and spelling. It does not correct and ‘judge’ students’ writing mistakes, but marks potential problem fields and provides students with balanced and concise feedback, tips, examples and links to informative websites.”

We are currently working on all these fronts with a view to improving the usability and efficiency of the tool before making it available as a stand-alone resource, or preferably as part of a versatile academic online writing package.

Bibliography

- Abel, Andrea (2010). Towards a systematic classification framework for dictionaries and CALL. In Granger, Sylviane & Paquot, Magali (eds.) *eLexicography in the 21st century. New challenges, new applications* (pp. 3-11). Louvain-la-Neuve: Université catholique de Louvain.
- Atkins, B.T. Sue & Rundell, Michael (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Bergenholtz, Henning & Tarp, Sven (2010). LSP lexicography or terminography? The lexicographer’s point of view. In Fuertes-Olivera, P.A. (ed.): 27-37.
- Biber, Doug, Johansson, Stig, Leech, Geoffrey, Conrad, Susan and Finegan, Edward (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Ltd.
- Bowker, Lynne (2010). The contribution of corpus linguistics to the development of specialized dictionaries for learners. In Fuertes-Olivera, P.A. (ed.): 155-168.

- Chang, Ji-Yeon (2014). The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL* 26(2), 243-259.
- Cobb, Tom (2003). Do corpus-based electronic dictionaries replace concordancers? In B. Morrison, G. Green, & G. Motteram (eds.). *Directions in CALL: Experience, experiments, evaluation* (pp. 179-206). Hong Kong: Polytechnic University.
- Coxhead, Averil (2000). A new Academic Word List. *TESOL Quarterly* 34(2): 213-238.
- de Schryver, Gilles-Maurice (2013). The concept of simultaneous feedback. In R.H. Gouws, U. Heid, W. Schweickard & H.E. Wiegand (eds.). *Dictionaries. An International Encyclopedia of Lexicography*. Supplementary volume: Recent Developments with Focus on Electronic and Computational Lexicography (Handbooks of Linguistics and Communication Science, HSK 5.4) (pp. 548–556). Berlin: Walter de Gruyter.
- De Wachter, Lieve, Verlinde, Serge, D’Hertefelt, Margot & Peeters, Geert (2014). How to deal with students’ writing problems? Process-oriented writing support with the digital Writing Aid Dutch. *Proceedings of COLING 2014*, Dublin, Ireland, August 23-29 2014, 20-24.
- Flowerdew, Lynne (2008). *Corpus-based Analyses of the Problem-Solution Pattern. A Phraseological Approach*. Amsterdam & Philadelphia: Benjamins.
- Fuertes-Olivera, Pedro A. (ed.) (2010). *Specialized Dictionaries for Learners*. Berlin & New York: Walter de Gruyter.
- Fuertes-Olivera, Pedro A. & Henning Bergenholtz (eds.) (2011). *e-Lexicography. The Internet, Digital Initiatives and Lexicography*. London & New York: Continuum.
- Gabel, S. (2001). Over-indulgence and under-representation in interlanguage: Reflections on the utilization of concordancers in self-directed foreign language learning. *Computer Assisted Language Learning* 14, 269-288.
- Gamper, Johann & Knapp, Judith. (2002), Adaptation in a Vocabulary Acquisition System. *KI - Zeitschrift Künstliche Intelligenz* 3(2): 27-30.
- Gilquin, Gaëtanelle, Granger, Sylviane & Paquot, Magali (2007) Learner corpora: the missing link in EAP pedagogy. *Journal of English for Academic Purposes* 6(4): 319-335.
- Gouws, Rufus H. (2010). The monolingual specialized dictionary for learners. In Fuertes-Olivera, P.A. (ed.): 55-68.
- Granger, Sylviane, Dagneaux, Estelle, Meunier, Fanny & Paquot, Magali (2009). *The International Corpus of Learner English. Handbook and CD-ROM. Version 2*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Granger, Sylviane & Paquot, Magali (eds.) (2012). *Electronic Lexicography*. Oxford: Oxford University Press.
- Hanks, Patrick (2013). *Lexical Analysis. Norms and Exploitations*. Cambridge & London: The MIT Press.
- Hardie, Andrew (2009). CQPweb – combining power, flexibility and usability in a corpus analysis tool. Paper presented at the 30th ICAME conference, Lancaster, 27-31 May 2009. Available from <http://www.lancs.ac.uk/staff/hardiea/cqpweb-paper.pdf> (last accessed 26 March 2015).
- Hartmann, R. R. K. (2005). Pure or hybrid? The development of mixed dictionary genres. *Facta Universitatis* 3(2). 193-208.
- Henry, O. (2000). Calloway’s code. In Lewis, M. (ed.) *Teaching Collocation. Further Developments in the Lexical Approach*. Thomson Heinle, 118-124.
- Jalali, Hassan (2014). Examining novices’ selection of lexical bundles: The case of EFL postgraduate students in applied linguistics. *Journal of Applied Linguistics and Language Research* 1(2), 1-11.
- Jónsson, J.H. (2009). Lexicographic description. An onomasiological approach on the basis of phraseology. In Nielsen, S. & Tarp, S. (eds.): 257-280.

- Katz, S. (1996). Distribution of common words and phrases in text and language modelling. *Natural Language Engineering* 2(1), 15-59.
- Kilgarriff, Adam (2009). Corpora in the classroom without scaring the students. Proceedings from the 18th International Symposium on English Teaching, Taipei. Available from [www.kilgarriff.co.uk/Publications/2009-K-ETA-Taiwan-scaring.doc_\(last accessed 26 March 2015\)](http://www.kilgarriff.co.uk/Publications/2009-K-ETA-Taiwan-scaring.doc_(last%20accessed%2026%20March%202015))
- Kilgarriff, Adam, Rychly, Pavel, Smrž, Pavel & Tugwell, David (2004). The Sketch Engine. In Williams, Geoffrey & Vessier, Sandra (eds.) *Euralex 2004 Proceedings*. Lorient: Université de Bretagne-Sud, 105-116.
- Kilgarriff, Adam, Husák, Miloš, McAdam, Katy, Rundell, Michael & Rychly, Pavel (2008). GDEX: Automatically finding good dictionary examples in a corpus. In Bernal, E. & DeCesaris, J. (eds.) *Euralex Proceedings 2008*. Leeuwarden: the Netherlands: Fryske Akademy, 372-379.
- Kübler, Natalie & Pecman, Mojca (2012). The ARTES bilingual LSP dictionary: From collocation to higher order phraseology. In Granger, S. & Paquot, M. (eds.): 187-209.
- Laufer, Batia (1992). Corpus-based versus lexicographer examples in comprehension and production of new words. In Tommola, H., Varantola, K., Salmi-Tolonen, T. & Schopp, J. (eds.) *EURALEX 1992 Proceedings*. Tampere: University of Tampere, 71-76. [reprinted in Fontenelle 2008]
- Laufer, Batia & Girsai, Nany (2008) Form-focused instruction in second language vocabulary learning: a case for contrastive analysis and translation. *Applied Linguistics* 29(4): 694-716.
- Laufer, Batia & Levitzky-Aviad, Tamar (2006). Examining the effectiveness of ‘bilingual dictionary plus’ – a dictionary for production in a foreign language. *International Journal of Lexicography* 19(2): 135-155.
- L’Homme, Marie-Claude (2010). Designing terminological dictionaries for learners based on lexical semantics: The representation of *actants*. In Fuertes-Olivera, Pedro A. (ed.) (2010a), 141-153.
- L’Homme, Marie-Claude, Robichaud, Benoît & Leroyer, Patrick (2012). Encoding collocations in DiCoInfo : From formal to user-friendly representations. In Granger, S. & Paquot, M. (eds.): 211-236.
- Lee, David & Swales, John (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1): 56-75.
- Müller-Spitzer, Carolin (2013). Textual structures in electronic dictionaries compared with printed dictionaries: A short survey. In R.H. Gouws, U. Heid, W. Schweickard & H.E. Wiegand (eds.). *Dictionaries. An international encyclopedia of lexicography. Supplementary volume: Recent developments with special focus on computational lexicography*. Berlin: Walter de Gruyter, 367–381.
- Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nielsen, Sandro (2013). The future of dictionaries, dictionaries of the future. In Jackson, Howard (ed.) *Bloomsbury Companion to Lexicography*. London & New York: Bloomsbury, 355-372.
- Nielsen, Sandro (2014). Example sentences in bilingual specialized dictionaries assisting communication in a foreign language. *Lexicos* 24, 198-213.
- Nielsen, Sandro & Sven Tarp (eds.) (2009). *Lexicography in the 21st Century. In honour of Henning Bergenholtz*. Amsterdam & Philadelphia: Benjamins.

- Oakes, Michael and Farrow, Malcolm (2007). Use of the chi-squared test to examine vocabulary differences in English language corpora representing seven different countries. *Literary and Linguistic Computing* 22 (1): 85–99.
- Paquot, Magali (2010). *Academic Vocabulary in Learner Writing*. London & New York: Continuum.
- Paquot, Magali (2012). The LEAD dictionary-cum-writing aid: An integrated dictionary and corpus tool. In Granger, S. & Paquot, M. (eds.): 163-185.
- Paquot, Magali & Granger, Sylviane (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics* 32: 130-149.
- Pearson, Jennifer (1998). *Terms in Context*. Amsterdam & Philadelphia: Benjamins.
- Pecman, Mojca (2008). Compilation, formalisation and presentation of bilingual phraseology: problems and possible solutions. In F. Meunier and S. Granger (eds) *Phraseology in Foreign Language Learning and Teaching* (pp. 203-222). Amsterdam and Philadelphia: Benjamins.
- Prinsloo, D.J. (2009). The role of corpora in future dictionaries. In Nielsen, S. & Tarp, S. (eds.): 181-206.
- Rodgers, Ornait, Chambers, Angela and Le Baron-Earle, Florence (2011) Corpora in the LSP classroom: A learner-centred corpus of French for biotechnologists. *International Journal of Corpus Linguistics* 16 (3): 391–411.
- Rundell, Michael (1998). Recent trends in English pedagogical lexicography. *International Journal of Lexicography* 11(4), 315-342.
- Scott, M. (2004). *WordSmith Tools 4*. Oxford: Oxford University Press.
- Scott, M. (2008). *WordSmith Tools 5*. Oxford: Oxford University Press.
- Svensén, Bo (2009). *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Tarp, S. (2009). Reflections on data access in lexicographic works. In Nielsen, S. & Tarp, S. (eds.): 43-62.
- Trap-Jensen, Lars (2013). Researching lexicographical practice. In Jackson, Howard (ed.) *Bloomsbury Companion to Lexicography*. London & New York: Bloomsbury, 35-47.
- Verlinde, Serge & Binon, Jean (2009). Pedagogical lexicography revisited. In Bergenholtz, Henning, Nielsen, Sandro & Tarp, Sven (eds.) *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Bern: Peter Lang, 69-89.
- Usami, Hiroko (2013). Using a learner corpus to improve distractors in multiple choice grammar questions. In Granger, Sylviane, Gilquin, Gaëtanelle & Meunier, Fanny (eds.) *Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead*. Louvain-la-Neuve: Presses universitaires de Louvain, 455-462.
- West, M. (1953). *A General Service List of English Words*. London: Longman.
- Widdowson, H. G. (2003). *Defining Issues in English Language Teaching*. Oxford: Oxford University Press.
- Yihua, Zhang & Qiping, Guo (2010). An ideal specialised lexicography for learners in China based on English-Chinese specialised dictionaries. In Fuertes-Olivera, P.A. (ed.): 171-192.