

Customising a general EAP dictionary to meet learner needs¹

Sylviane Granger & Magali Paquot
Centre for English Corpus Linguistics
Université catholique de Louvain

Abstract

In this article, we describe the *Louvain EAP Dictionary (LEAD)*, a web-based English for Academic Purposes dictionary-cum-writing aid for non-native writers. The *LEAD* dictionary contains a rich description of non-technical words that express key functions in academic discourse (such as contrast, exemplification or cause and effect), with particular focus on their phraseology (collocations and recurrent phrases). The dictionary allows for both semasiological and onomasiological access. Its main originality is its customisability: the content is automatically adapted to users' needs in terms of discipline and mother tongue background.

1. Introduction

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 their MA/PhD 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). And crucially, a highly revealing correlation has been found between national output of international scientific publications and national English proficiency level (Man et al 2004).

There is obviously an urgent need to design tools that meet the needs of non-native speakers, be they English as a Foreign/Second Language (EFL/ESL) students or young researchers. In order to help them master the vocabulary needed in academic settings, a number of word lists have been compiled. The *Academic Word List (AWL)* (Coxhead 2000), for example, consists of 570 word families which have wide range and reasonable frequency of occurrence in a large corpus of academic texts, but are not among the 2,000 most frequent English words. Hyland and Tse (2007: 238), however, note considerable differences in the use of academic vocabulary across disciplines and call into question the very existence of a general academic vocabulary. They argue that “all disciplines shape words for their own uses” (ibid: 240), as

¹ This article was originally published in Granger & Paquot (eds) (2010). *eLexicography in the 21st century: New challenges, new applications*. Proceedings of ELEX2009. Cahiers du CENTAL N°7, 87-96. Louvain-la-Neuve, Presses universitaires de Louvain. <http://www.i6doc.com/en/collections/cental/>

demonstrated by their clear preferences for particular meanings and collocations. The noun *strategy*, for example, has different preferred associations across disciplines (e.g. *marketing strategy* in business, *learning strategy* in applied linguistics and *coping strategy* in sociology).

While this register variability is a source of difficulty for native and non-native writers alike, non-native writers have a harder time as they also tend to transfer into English words and phrases that are typical of academic discourse in their first language (L1). This tendency induces a wide range of difficulties: semantic errors (e.g. French learners' often wrongly use 'indeed' in the sense of 'en effet'), lexico-grammatical errors (e.g. French learners will write 'discuss about' mapped on French 'discuter de'), awkward collocations (e.g. *put forward a conclusion; a conclusion must be gathered*), and rhetorical or stylistic infelicities (e.g. French learners tend to overuse sequences with 1st person plural imperative – 'let us examine', 'let us take the example of' – which reflect different rhetorical conventions in French academic writing) (see Gilquin *et al.* 2007; Paquot 2008, 2010; Granger and Paquot 2009b).

In this article, we describe the rationale behind an electronic tool which aims to meet some of the difficulties that beset EFL writers in academic settings. Section 2 describes academic vocabulary and focuses more particularly on the difficulties it poses to non-native writers. Section 3 discusses a key lexicographic development afforded by the electronic medium, i.e. dictionary customisation. Section 4 introduces the *Louvain EAP Dictionary (LEAD)*, a web-based EAP dictionary-cum-writing aid tool for non-native writers and Section 5 concludes.

2. Academic vocabulary, phraseology and learner writing

As demonstrated by Howarth's (1996/1998) study of verb + noun collocations in a corpus of social science texts, a large proportion of non-technical collocations in academic writing consist of a verb in a figurative sense and an abstract noun (e.g. *adopt + approach, reach + conclusion, obtain + result*). Howarth suggests that these collocations are an essential part of the procedural vocabulary of academic discourse. The author further argues that it is "not idioms that learners need for effective communication" (Howarth 1996: 156), at least in academic settings. Learners need the lexical means that will allow them to conform to "the native stylistic norms for a particular register", which "entails not only making appropriate grammatical and lexical choices but also selecting conventional [multi-word units] to an appropriate extent" (Howarth 1998: 186). Recent studies have shown that the highly conventionalised nature of academic discourse stems largely from 'lexical extensions' of a set of academic words such as *conclusion, issue, claim* or *argue*. These words acquire their organisational or rhetorical function in specific word combinations that are essentially semantically and syntactically compositional (e.g. *as discussed below, an example of ... is ..., the aim of this study is to..., it has been suggested, final outcome, direct result*) (e.g. Curado Fuentes 2001; Pecman 2008; Siepmann 2005). Most of these studies have also highlighted the extent to which there is commonality across academic genres and disciplines and thereby brought support to Gledhill's view that "there is a shared scientific voice or 'phraseological accent' which leads much technical writing to polarise around a number of stock phrases" (Gledhill 2000:204).

Learner corpus research has shown that there is little variation in the way EFL learners organise their papers and that they make scant use of lexico-grammatical patterns typical of academic discourse. In her study of verb + noun combinations in German learner writing, Nesselhauf (2005) notes that “the unavailability of pragmatic chunks for the learners (...) appears to be the underlying reason for a number of deviant collocations which are used to structure the body of the essay, (to introduce examples, for instance)” (2005: 141). Chunks such as *Only have a look at*, *If you have a look at*, *Let us have a look at*, *A first argument I want to name for this* are good illustrations of this kind of pragmatic failure. De Cock’s (2003) study of prefabs in learner writing illustrates another aspect of stylistic deficiency, viz. learners’ tendency to overuse a whole set of informal word sequences such as *and so*, *I think* and *there are a lot of* that confers a speech-like quality to their writing. She shows that learners are generally unaware of “the more common, less salient and frequently used L2 multi-word building blocks” (De Cock 2003: 65).

Learners’ use of phraseological patterns is also characterised by erroneous collocations and first language influence. Nesselhauf (2005), for example, shows that the most frequent types of error in verb + noun combinations produced by German EFL learners involve the erroneous choice of verb and in 56% of cases, are likely to result from transfer from the learner’s L1. Among the nouns that are most often used with deviant verbs are *action*, *aim*, *attitude*, *problem*, *question*, *statement*, *step* and *conclusion*, which fulfil key rhetorical functions in academic discourse. De Cock (2003) also showed that French learners (1) misuse English sequences that have a French congruent form which may be used differently, e.g. *on the contrary/au contraire*; (2) underuse multi-word units which have no literal L1 counterpart, e.g. *sort of*; and (3) use L1-induced idiosyncratic combinations, e.g. *according to me*.

3. Dictionary customisation

Dictionaries have traditionally been designed as “one-size-fits-all package[s]” (Rundell 2007: 50). Learners’ dictionaries, in particular, target a generic learner and claim to cater for their supposedly similar needs. One of the reasons behind this format is purely economic: it stems from the wish to reach the widest possible market with one single product and thereby maximise profits. This position is no longer defensible today. Recent research has illustrated the clear necessity to adapt dictionaries to users’ needs and technological advances have simultaneously put this development within the reach of dictionary producers. For many specialists, customisation is one of the main challenges of present-day lexicography. Tarp (2009a: 25), for example, points out that “lexicographic needs are not abstract needs, but are always related to specific types of users who find themselves in a specific type of social situation”. Put differently, “users in general never need information in general” (Tarp 2009b: 46). One way of implementing customisation is to replace the static data in electronic dictionaries “by articles containing dynamic data which are, so to say, unique for each search related to a specific type of user in a specific type of user situation” (Tarp (2009a: 29). It must be admitted, however, that there have been but few concrete achievements to date in spite of the fact that the idea of dictionary customisation has been around for quite some time (cf.

Atkins 2002; Varantola 2002; De Schryver 2003). As noted by Sobkowiak (2002), only superficial elements of customisation have been integrated:

[O]nly the rather superficial customizing options are offered, such as, for example: (a) ignoring certain elements of the entry (micro)structure for screen display (e.g. phonetic transcription) or in full-text search (e.g. example sentences), (b) hiding certain word categories (e.g. compounds), (c) changing font size, style and colours, (d) manipulating toolbars, and the like. (Sobkowiak 2002)

As the difficulties posed by academic English have proved to vary according to users' profile and context of use, in particular their mother tongue background and the discipline they write in, it is a field that would greatly benefit from customisation. It was this that prompted us to embark on the development of a customisable EAP dictionary which will be briefly outlined in the following section.

4. The Louvain EAP Dictionary

The *Louvain EAP Dictionary (LEAD)* is a corpus-based tool: 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 a number of home-made discipline-specific corpora) and EFL learner corpora representing a wide range of L1 populations. As shown in Figure 1, the dictionary contains a rich description of 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**. The dictionary relies on a relational MySQL database, 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 it makes full use of the capabilities afforded by the electronic medium in terms of multiplicity of access modes (Sobkowiak 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 and Levitzky-Aviad 2006) as users who have selected a particular mother tongue background can search lexical entries via their translations into that language.

The screenshot shows the 'The Louvain EAP dictionary' interface. At the top, there is a search bar with a 'Search' button and a dropdown menu showing 'Selected discipline: Business; selected mother tongue: French (change settings)'. Below the search bar are navigation tabs: 'Home', 'Word search', 'Search by function', 'Search by translation', and 'Corpus search'. The main content area displays the entry for 'example (n.)'. It includes a definition: 'a typical member of a group of things: Salisbury Cathedral is a classic example of English Gothic architecture.' and another definition: 'a way of showing someone how something is used to help them understand: The following examples show how this equation works in practice.' To the right, there are sections for 'Phrases' (example of, X is an example of Y, An example of Y is X) and 'Collocations' (Adj + example: clear, extreme, fine, good, notable, obvious, outstanding, perfect, prime, shining, striking, typical; example + V: demonstrate, illustrate, include, indicate, show, suggest). At the bottom, a highlighted box contains the sentence: 'In practice, the credit multiplier in the United Kingdom is not as large as the above example suggests because of leakages.'

Figure 1. The Louvain EAP dictionary

Before using the dictionary, users 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 2). This stage conforms to Tarp's (2009b: 48) suggestion "to prepare a preliminary interactive phase where the lexicographic tool helps the users to identify and specify their concrete needs before being guided to the corresponding data". Discipline-oriented customisation is currently being implemented in the selection of **examples** of collocations and phrases. The characteristics of good dictionary examples have been clearly identified by Atkins and 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* is illustrated by example (1) when the user has selected business as the target discipline and example (2) when medicine is the target, thereby adhering 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".

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.

As shown in Figure 1, clicking on a specific collocate or phrase (e.g. *suggest*) displays a discipline-specific example of the phraseological pattern (*as the above example suggests*) in a box at the bottom of the lexical entry.

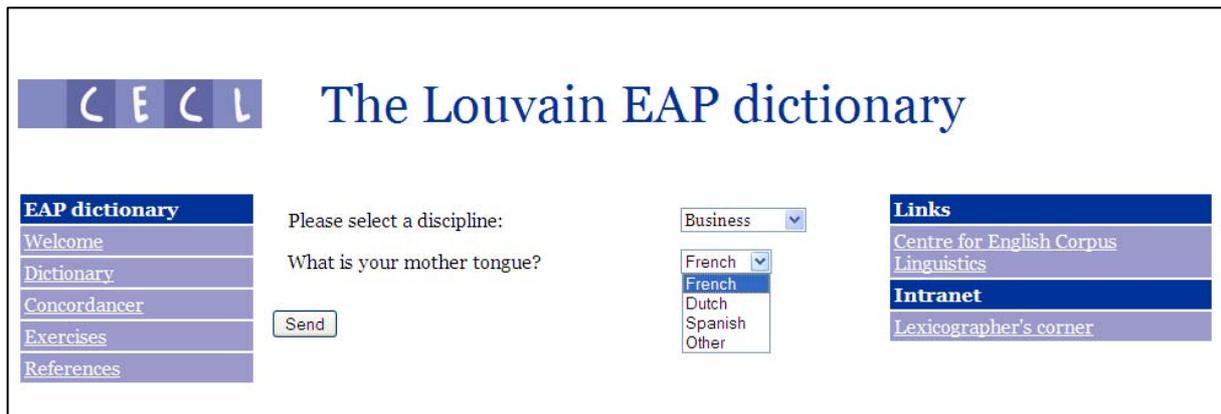


Figure 2. Customising the Louvain EAP dictionary

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, warnings about common translation mistakes are also included, such as the erroneous translation of the French ‘prétendre’ by its false friend ‘pretend’ in English. We are currently focusing on French as an 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 of the *Varieties of English for Specific Purposes dAtabase (VESPA)*, a new learner corpus, currently being developed at the Centre for English Corpus Linguistics in collaboration with several international partners. The corpus contains L2 texts from a wide range of L1 backgrounds (currently French, Spanish, Swedish, and Polish), 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). Errors and difficulties found in the writing of a wide range of learner populations are dealt with in generic error notes that are displayed irrespective of the L1 background selected by the user (cf. Figure 3). Errors found exclusively in the writing of French learners are described in notes that only show up if French is selected as L1 background. Thus, the lexical entry for “according to” includes an error note that draws French users’ attention to the erroneous translation of French “selon moi” by English “according to me”.

Onomasiological access to the dictionary is via 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 4). Selecting one of these functions provides the user with a list of lexical items, categorised according to part-of-speech (nouns, verbs, adjectives, adverbs, conjunctions and prepositions), typically used to serve this function in academic texts. One of the main advantages of this access mode is that it suggests alternatives and thereby helps users enlarge their academic repertoire. Words are

currently sorted alphabetically but they could be sorted by frequency of occurrence in the discipline-specific corpora. Each word is clickable and users can get access to its full lexical entry.

namely (*adv.*)

used to go into more detail about or identify something you have just mentioned:

- *By the early 19th century, England still only had two universities, the same two which had been there since the thirteenth century, **namely** Oxford and Cambridge.*
- *The D-Day beaches in Normandy are still known by the code names given to them during wartime, **namely** Utah, Omaha, Gold, Juno and Sword.*
- *As the early universities took shape, an important distinguishing feature started to emerge, **namely** the manner and extent to which the institution would engage in teaching the professions, as opposed to the liberal arts.*

Error note

Don't use **namely** to introduce examples. Use **such as**:

- *Such a situation is due to the fact that people prefer easier entertainment, **such as** watching television or playing computer games.*

Figure 3. An example of a generic error note

C E C L The Louvain EAP dictionary

Add information Search

Function Search by translation Corpus search

Describe differences

Adverbs	Conjunctions	Prepositions
by comparison	whereas	by comparison with
by contrast	while	contrary to
conversely		in comparison with
in comparison		in contrast to
in contrast		in contrast with
on the other hand		unlike
		versus

Figure 4. Onomasiological access

The *LEAD* dictionary is designed as an integrated tool where the actual dictionary part is linked up to other language resources (in particular, discipline-specific corpora and a corpus handling tool). We are currently turning the *LEAD* into a dictionary-cum-CALL resource (Abel this volume) by adding exercises targeting learners' attested difficulties.

5. Conclusion

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 academic functions such as contrasting or reporting. 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. Parallel to these findings, recent research has put needs analysis at the heart of both EAP course design and teaching (e.g. Jordan 1997; Hyland 2002) and lexicography (Tarp 2008). The *Louvain EAP Dictionary* is an attempt to implement these findings in a customisable web-based tool. 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). The dictionary is currently a stand-alone product but it could – and ideally should – be integrated into a general dictionary and/or a suite of teaching and learning tools.

References

- ATKINS, B.T.S. (2002). Bilingual dictionaries - Past, present and future. In M.-H. Corréard (ed.) *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*. Euralex, 1-29.
- ATKINS, S. and RUNDALL, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- COXHEAD, A. (2000). A new Academic Word List. *TESOL Quarterly*, 34(2): 213-238.
- CURADO FUENTES, A. (2001). Lexical behaviour in academic and technical corpora: Implications for ESP development. *Language Learning & Technology*, 5(3): 106-129.
- DE COCK, S. (2003). *Recurrent sequences of words in native speaker and advanced learner spoken and written English: a corpus-driven approach*. Unpublished PhD thesis. Louvain-la-Neuve: Université catholique de Louvain.
- DE SCHRYVER, G.-M. (2003). Lexicographers' dreams in the electronic-dictionary age. *International Journal of Lexicography*, 16(2): 143-199.
- GILQUIN, G., GRANGER, S. and PAQUOT, M. (2007). Learner corpora: the missing link in EAP pedagogy. In P. Thompson (ed.). *Corpus-based EAP Pedagogy*. Special issue of *Journal of English for Academic Purposes*, 6(4): 319-335.
- GLEDHILL C. (2000). *Collocations in Science Writing*. Language in Performance 22. Tuebingen: Gunter Narr Verlag.

- GRANGER, S., DAGNEAUX, E., MEUNIER, F. and PAQUOT, M. (2009). *The International Corpus of Learner English. Handbook and CD-ROM. Version 2*. Louvain-la-Neuve: Presses universitaires de Louvain. Available from <http://www.i6doc.com>
- GRANGER, S. and PAQUOT, M. (2008). From dictionary to phrasebook? In E. Bernal and J. DeCesaris (eds) *Proceedings of the XIII EURALEX International Congress*. Barcelona, Spain, 15-19 July 2008, 1345-1355.
- GRANGER, S. and PAQUOT, M. (2009a). In search of General Academic English: A corpus-driven study. *Proceedings of the International Conference on L.S.P.: 'Options and Practices of LSP Practitioners'*, 7-8 February 2009, Heraklion, Crete. Available from <http://cecl.fltr.ucl.ac.be/publications.html>.
- GRANGER, S. and PAQUOT, M. (2009b). Lexical verbs in academic discourse: A corpus-driven study of learner use. In M. Charles, S. Hunston and D. Pecorari (eds) *At the Interface of Corpus and Discourse: Analysing Academic Discourses*. London: Continuum, 193-214.
- HOWARTH, P. (1996). *Phraseology in English Academic Writing: Some Implications for language learning and dictionary making*. Tübingen: Max Niemeyer Verlag.
- HOWARTH, P. (1998). The phraseology of learners' academic writing. In A.P. Cowie (ed.), *Phraseology: theory, analysis and applications*. Oxford: Oxford University Press, 161-186.
- HYLAND, K. (2002). Specificity revisited: how far should we go now? *English for Specific Purposes*, 21:385-395.
- HYLAND, K. and TSE, P. (2007). Metadiscourse in academic writing: a reappraisal. *Applied Linguistics*, 25(2): 156-177.
- LAUFER, B. and LEVITZKY-AVIAD, T. (2006). Examining the effectiveness of 'bilingual dictionary plus' – a dictionary for production in a foreign language. *International Journal of Lexicography*, 19(2): 135-155.
- MAN, J.P., WEINKAUF, J.G., TSANG, M. and SIN, D.D. (2004). Why do some countries publish more than others? An international comparison of research funding, English proficiency and publication output in highly ranked general medical journals. *European Journal of Epidemiology*, 19: 811-817.
- MOON, R. (2008). Dictionaries and collocation. In S. Granger and F. Meunier (eds) *Phraseology. An Interdisciplinary Perspective*. Amsterdam: Benjamins, 313-336.
- MUNGRA, P. and WEBBER, P. (2010). Peer review process in medical research publications. Language and content comments. *English for Specific Purposes*, 29(1) : 43-53.
- NESSELHAUF N. (2005). *Collocations in a learner corpus*. Amsterdam: Benjamins.
- PAQUOT, M. (2008). Exemplification in learner writing: A cross-linguistic perspective. In F. Meunier and S. Granger (eds) *Phraseology in Foreign Language Learning and Teaching*. Amsterdam and Philadelphia: Benjamins, 101-119.
- PAQUOT, M. (2010). *Academic Vocabulary in Learner Writing*. London & New York: Continuum.
- PAJZS, J. (2009). On the possibility of creating multifunctional lexicographical databases. In H. Bergenholtz, S. Nielsen and S. Tarp (eds) *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Bern: Peter Lang, 327-354.
- PECMAN, M. (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*. Amsterdam and Philadelphia: Benjamins, 203-222.
- RUNDELL, M. (2007). The dictionary of the future. In S. Granger (ed.) *Optimizing the role of language in technology-enhanced learning*. Proceedings of the expert workshop organized in Louvain-la-Neuve (Belgium), 4-5 October 2007, 49-51.

<http://jupiter.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/Downloads/GRANGER-SYLVIANE-2007.pdf>

- SIEPMANN, D. (2005). *Discourse markers across languages: a contrastive study of second-level discourse markers in native and non-native text with implications for general and pedagogic lexicography*. London and New York: Routledge.
- SOBKOWIAK, W. (2002). What can be but is not (and why) in learners' MRDs. *Teaching English with Technology*, 2(3). Available at http://www.iatefl.org.pl/call/j_article9.htm.
- TARP, S. (2008). *Lexicography in the Borderland between Knowledge and Non-Knowledge*. Lexicographica Series Maior 134. Tübingen: Niemeyer.
- TARP, S. (2009a). Beyond lexicography: New visions and challenges in the information age. In H. Bergenholtz, S. Nielsen and S. Tarp (eds) *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Bern: Peter Lang, 17-32.
- TARP, S. (2009b). Reflections on data access in lexicographic works. In S. Nielsen and S. Tarp (eds). *Lexicography in the 21st Century*. Amsterdam / Philadelphia: Benjamins, 43-62.
- VARANTOLA, K. (2002). Use and usability of dictionaries: common sense and context sensitivity? In M.-H. Corréard (ed.) *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*. Euralex, 30-44.