

Has Lexicography Reaped the Full Benefit of the (Learner) Corpus Revolution?

Sylviane Granger

Centre for English Corpus Linguistics, Université catholique de Louvain

E-mail: sylviane.granger@uclouvain.be

Abstract

In 1992 Rundell and Stock wrote an extended three-part article on the “corpus revolution”, in which they describe the rise of corpora and their impact on lexicography. Rundell (2008) revisited the topic and focused on the arrival of the web, which triggered a second stage in the corpus revolution. The purpose of my presentation is to look at some aspects of the current lexicographic scene and assess whether the corpus revolution has really fulfilled its promise. I will show that, while this is largely true of monolingual learners’ dictionaries (especially in the case of English), the situation is much less favourable when it comes to general bilingual dictionaries, a particularly worrying fact given that bilingual dictionaries have been proved to be learners’ favourite reference tool. The lack of representative translation corpora is partly responsible for this failure to keep pace. However, I will show that the judicious use of currently available monolingual and bilingual corpus resources, limited though they may be, can already bring about substantial improvements to bilingual dictionaries, in particular to one of their weakest aspects: their phraseological coverage. I will also highlight the value of adding learner corpus data to the lexicographer’s monolingual and bilingual corpus base and illustrate the benefit with the *Louvain English for Academic Purposes Dictionary*, a customisable web-based tool designed to help non-native speakers of English write academic texts.

Keywords: phraseology, learner corpora, bilingual corpora, bilingual dictionaries

1 The Corpus Revolution

In their three-part 1992 article entitled *The corpus revolution* Michael Rundell & Penny Stock described the “exciting and unnerving changes” to lexicography brought about by the emergence of a new type of lexicographic evidence, namely large collections of texts in electronic format which can be stored and processed by computers. In *The corpus revolution revisited*, written in 2008, Michael Rundell described a second stage in the corpus revolution triggered by the accessibility of much larger quantities of language data from the web. Ten years on it seems timely to take stock of the situation, highlighting both successes and weaknesses and pointing to possible avenues of improvement. In particular, as foreign language learning and teaching was at the heart of A.S. Hornby’s work, it seems appropriate, in a lecture in his honour, to assess the impact of learner corpora, a corpus resource that emerged in the early 1990s and which I, perhaps somewhat prematurely, presented as heralding another revolution, this time in the wider field of applied linguistics (Granger 1994). A fully comprehensive survey of the field is beyond the scope of this lecture and in any case exceeds my own level of expertise. My focus will be on general (i.e. non-technical) bilingual and pedagogical lexicography.

There are many different types of corpus and each has its own set of lexicographic uses. The corpora that can truly be said to have revolutionised lexicography are monolingual corpora representing native/expert speaker language. In today’s digital world, vast amounts of electronic text of this nature can be collected fairly easily, thereby ensuring that the criteria of size and balance are met. Nowhere is this revolution more in evidence than in monolingual learners’ dictionaries (MLDs) and

more particularly, in their “concern for describing and explaining phraseology” (Rundell 1998:318). Although the pioneering work of Hornby and Palmer paved the way for a more phrasal approach to lexis, it is the advent of the computer corpus that has allowed it truly to materialise. Sinclair’s innovative corpus-driven approach to lexicography, in particular, led to a formidable extension of the field of phraseology, whose scope came to embrace an increasingly large and diversified set of units, including collocations, colligations, lexical bundles and semantic prosodies.

Although the greatest contribution to monolingual learners’ dictionaries is that of native corpus data, a more limited, but highly promising, impact has been made by learner corpus data. In fact, MLDs were the first language resource to make use of this type of corpus. Based on an idea initially put forward by Maingay & Rundell (1987), Longman included error notes based on the *Longman Learner Corpus* in the third edition of the *Longman Dictionary of Contemporary English* (1995). Two other publishing houses – Cambridge and Macmillan – followed suit. In the *Macmillan English Dictionary for Advanced Learners* (2007), the notes, which took the form of ‘Get it right’ boxes, resulted from a close collaboration between Macmillan and the Centre for English Corpus Linguistics at the University of Louvain. Close investigation of the *International Corpus of Learner English* (Granger et al. 2009) resulted in extended error notes designed to warn users against common pitfalls (Rundell & Granger 2007). This was a valuable first step but as I will show in this article, it is possible and desirable to make a much more ambitious use of learner corpus data.

Fully justified as a description of the development of English monolingual dictionaries, the term ‘revolution’ is much less appropriate in the context of bilingual dictionaries. Although there can be no doubt about the potential usefulness of translation corpora in bilingual lexicography, their actual use is still very limited: “One area where one might expect such corpora to be widely used is bilingual lexicography, but in fact such corpora have not been exploited significantly in dictionary compilation – unlike monolingual lexicography, where it would be unthinkable today not to use single-language corpora” (Salkie 2008). As a result, “[t]he extraordinary range of lexical and grammatical information they [MLDs] include is rarely even approached by the best bilingual dictionaries available” (Rundell 1999). Twenty years on Rundell’s observation remains largely true. The main reason is the unavailability of representative, balanced translation corpora. Many translation corpora have been compiled by academic research teams but they are rarely publicly available. Those that are in the public domain often represent specialised registers of English (e.g. parliamentary debates in *EuroParl*) and fail to differentiate between source and target language, a serious weakness in view of the possibility of source text influence on translated texts. Other factors play a part in the – hopefully temporary – failure of corpus-based bilingual lexicography to keep pace, notably the lack of adequate tools and training to facilitate bilingual corpus exploration by lexicographers. This is not to say that the corpus revolution has entirely passed bilingual lexicography by. Some advances have been made but there is still ample room for improvement. It is to be hoped that the development of user-friendly corpus tools, in particular bilingual functionalities such as those included in *Sketch Engine* (Matuška 2018), will lead to a more systematic use of corpus data. This is particularly desirable given that several studies have shown that bilingual dictionaries are learners’ favourite reference tool. Nesi (2014: 38), for example, notes that “[a]lthough it is commonly believed that monolingual dictionaries are superior to bilingual dictionaries in terms of their usefulness as language learning tools, attitude and ownership surveys have found that learners generally prefer to use bilingual dictionaries”. This is because learners, even at an advanced stage of acquisition, still regularly think in their mother tongue (L1) and, when the target word escapes them, have no alternative but to turn to bilingual dictionaries. Even bilingualised dictionaries will not help as they are essentially monolingual dictionaries supplemented with very succinct bilingual information.

Clearly then, while the term ‘corpus revolution’ is certainly not undeserved, there is still scope for more intensive and diversified use of corpus data in lexicography. In the following sections I will

describe two directions that can be taken in order to help achieve that objective: (1) the combined use of monolingual and translation corpora to improve the phraseological coverage of bilingual dictionaries and (2) more extensive use of learner corpus data and its integration into web-based lexical environments.

2 Phraseology in General Bilingual Dictionaries

For a long time the phraseological coverage of bilingual dictionaries was restricted to semantically non-compositional, often figurative, units such as idioms and proverbs. The situation has fortunately improved in recent years and many bilingual dictionaries now also contain rich descriptions of collocations. However, the phraseological information provided in bilingual dictionaries is still generally felt to be insufficient (Farina 2009, Gouws 2010, Mogorrón Huerta 2011, Xia 2015). A large proportion of the prefabricated units uncovered by corpus methods, especially those that are semantically compositional, is left on the sidelines. This applies, for example, to Biber et al.'s (1999 "lexical bundles", i.e. recurrent contiguous word sequences that can easily be identified by the n-gram method, which extracts all sequences of a specific number of words (e.g. 4-word sequences) or in a given range (3- to 6-grams). The n-gram method has been commonly used to extract specialised bilingual terminology, but seems to have been underused in general bilingual lexicography. This is unfortunate as these sequences often play a key role in discourse, as textual organisers or stance markers. In addition, many are polyfunctional, which makes them particularly difficult to translate and justifies their inclusion in bilingual dictionaries. Examples of these sequences in French are: *vous n'êtes pas sans savoir que* (you are no doubt aware that, you certainly know that; lit. you are not without knowing that), *en matière de* (regarding, with regard to; lit. in matter of), *ces derniers temps* (lately, recently; lit. these last times), *il n'en reste pas moins que* (the fact remains that; lit. it no less remains that) and *en provenance de* (from; lit. in provenance from).

In view of the importance of these units and the difficulty they represent, in particular for encoding purposes, it seems worthwhile to check for their presence in general bilingual dictionaries, identify the status they are given in the micro- and macro-structure of the dictionary and assess the quality of the translations provided. This has been the purpose of recent research (Granger & Lefer 2012, 2013, 2016, Granger in press) focused on French-English general bilingual dictionaries. The methodology used involved two types of corpus: monolingual corpora to extract lexical bundles and translation corpora to extract authentic translations of the bundles. The corpus-extracted phrases and their translations were subsequently set against the descriptions provided in electronic French-English dictionaries.

In spite of its restricted scope, the research turned out to be highly instructive. First, it showed that it is possible to extract a sizeable number of useful phrases even from a small monolingual corpus of French. In Granger & Lefer (2012) we were able to identify 425 dictionary-worthy sequences of 2 to 5 words from a 1-million word corpus of French. A comparison with two French-English electronic dictionaries (*Le Grand Robert & Collins v2* and *Hachette Oxford*) revealed that 12-15% of these sequences were absent from the French-English part of the dictionaries. A follow-up study centred on the phraseology of high-frequency adverbs such as *encore* in French or *yet* in English (Granger & Lefer 2013) suggests that dictionaries tend to include phrases that are more typical of speech (*et puis quoi encore!* What next!), while the corpus brings to light phrases typical of writing, many of them with linking functions (*l'Italie, l'Espagne ou encore la France*: Italy, Spain or France). This difference is most probably due to the fact that lexicographers still rely mainly or exclusively on introspection and quite naturally tend to think of interactive markers that are more emotionally loaded than the fairly inconspicuous cohesive markers typical of writing, which can only be brought to consciousness

by corpus methods. In this area, especially given the absence of spoken translation corpora, there is clear complementarity between introspection-based and corpus-derived insights.

A second important result is that when present, a large proportion (c. 50%) of the multiword units (MWUs) are included as subentries. Not only does this status make them more difficult to access, even in electronic dictionaries, but more importantly, it accounts for the very cursory treatment they are given, often limited to one translation and/or one example sentence. In the case of a quarter of the units the situation is even worse as they appear only in an example sentence. The proportion of sequences given head-entry status was found to be very limited (4% to 13%), thereby confirming Jackson's (2013) observation that "[m]ultiword expressions [...] are most of the time treated in the microstructure although their fixed status (when they have one, which is not always the case) could justify their presence in the list of headwords". Headword status would be fully justified for word-like units which are in effect what Palmer (1917) called "accidents of graphic continuity", i.e. fixed units that are to all intents and purposes words. This would go some way towards fulfilling Sinclair's (2010: 37) general call to award headphrase status to MWUs: "The evidence from corpora adds up to a strong case for extending the treatment of multi-word units of meaning— a much wider concept than idiom— and giving them the same status as the usual headwords". Headphrase status may not be realistic for all types of phraseological units but it is certainly desirable for word-like units such as multiword prepositions (*due to, instead of, in accordance with*), multiword conjunctions (*even if, provided that, insofar as*) and multiword adverbs (*of course, on the other hand, in my opinion*), which are very numerous.

A third finding, already established by several other corpus-based studies, is that the translations provided by bilingual dictionaries and those extracted from translation corpora are very different. A first observation in this connection is that many translations suggested by bilingual dictionaries are infrequently attested in translation corpora. This is partly due to lexicographers' tendency to translate MWUs literally, a tendency that is reinforced by the fact that literal translations are often possible. For example, *de la même manière/façon* can – in some contexts – be translated by *in the same way*. However, corpus analysis shows that this literal translation, which is promoted in bilingual dictionaries, is only appropriate when the French phrase is used as an adverb of manner. When it is used as a linking adverb, the most usual translations are *similarly* and *likewise*. This example illustrates another typical characteristic of dictionary translations: the tendency to translate an MWU by another MWU, rather than by a single-word equivalent. Bilingual dictionaries also contain ample evidence of what Granger & Lefer (2012) refer to as 'categorical bias', i.e. lexicographers' tendency to translate a given source item exclusively into a word of the same grammatical category in the target language (to translate an adjective into an adjective, an adverb into an adverb, etc.). By contrast, corpus translations often involve changes in grammatical category, resulting in a host of "nice surprises" (Salkie 2008), i.e. excellent translations unattested in bilingual dictionaries. This difference raises challenging questions on the types of equivalence that bilingual dictionaries should include: systemic or textual equivalence (Adamska-Sałaciak 2010). As argued in Granger & Lefer (2016), in a production-oriented dictionary it seems legitimate to give textual equivalence a more prominent place than is currently the case.

Translation corpora have undeniable lexicographic value. However, one important caveat regarding their use must be mentioned: translation corpora are not error-free. Even when the texts have been translated by professional translators, they may contain numerous examples of calques and infelicitous translations. This is especially common in the case of discourse-oriented phrases whose pre-fabricated nature may pass unnoticed in view of their semantic compositionality and the possibility of literal translation. The role of lexicographers remains essential, since weeding out infelicitous or erroneous translations and selecting the best candidates for inclusion in the dictionary requires a level of expertise that only human experts can provide.

3 More Extensive Use of Learner Corpus Data

Learner corpora are a relatively new resource in the constellation of corpora that has been built over the years. One of their main benefits is that they make it possible to draw up catalogues of learners' difficulties at any stage in the learning process. These difficulties can be assessed in terms of over- and underuse of particular words, phrases or structures, and/or outright errors. As stated in the introduction, several publishing houses have availed themselves of this type of information as a basis for inserting error notes in MLDs. However, work on this front has been at a standstill for a number of years now. One of the reasons, besides possible financial motives, may well be that lexical errors tend to be L1-specific, and only errors shared by a large number of L1 populations are suitable for inclusion in generic dictionaries like MLDs. The next step forward is clearly to customise MLDs, partly in order to cater for the specific needs of different L1 populations. This trend towards customisation has been advocated by numerous authors. Rundell (2007: 50) foresees the end of "the current globally-marketed one-size-fits-all package" and sees future reference materials "as a set of components which customers can mix and match according to their needs". Individualisation also lies at the heart of function theory which describes the role of dictionaries as that of "meeting the specific types of information which a specific type of users may have in a specific type of situation" (Tarp 2009: 47).

Web-based dictionaries have a high degree of flexibility in the type and quantity of information that can be presented to users and are therefore the ideal resource to implement customisation. Another key advantage is that they allow for a high degree of interconnectivity: they can easily be integrated into wider environments featuring components such as writing aids, exercises and direct corpus access, to name just a few. Although the "one-stop shopping" dreamed of by Bowker (2010: 166-7) is probably unrealistic, at least in the foreseeable future, every effort needs to be made to reduce the number of tools that users have to open when performing a particular task, such as writing. A few tools that combine the functions of dictionary and learning/writing aid, have been developed or are currently under development (see, for example, Verlinde & Peeters 2012, García-Salido et al. 2018). In the following lines I describe the *Louvain English for Academic Purposes Dictionary* (LEAD)¹ (Granger & Paquot 2015), a hybrid tool designed to help non-native writers of English with writing academic texts.

The launch of the LEAD initiative was prompted by a keen awareness that an increasing number of non-native students and researchers were having to write academic texts in English but could not find any tools that met their specific writing needs. At a higher intermediate/advanced proficiency level, students and researchers usually have a good mastery of grammar and an extensive vocabulary, including domain-specific vocabulary in the case of ESP students, but fail to produce texts that conform to the typical academic style. In particular, while such writers may know the meaning of a large number of words or phrases typically used in academic texts, whatever the discipline (such as *to conclude, however, research*), they are not always aware of their distinct academic "priming" (Hoey 2005) in terms of grammar (passive or active voice), position (initial, medial or final), style (formal or informal) and, most importantly, preferred lexico-grammatical patterning, i.e. collocations and lexical bundles. For example, learners regularly use the erroneous collocation *to make research* rather than the correct *to conduct, carry out, undertake research*. They also tend to use atypical extended sequences to convey important discourse functions such as concluding, comparing or illustrating. For example, they may conclude their essays with sequences such as *I would conclude by saying that* or *I want to conclude saying that* rather than the more typical *it can/may be concluded that* or *we may conclude that*. The LEAD provides collocations and lexical bundles for over 1,200 academic words. We made use of a corpus of expert writing (academic part of the *British National Corpus*) to extract typical patterning, and of a large learner corpus, the *International Corpus of Learner English*

¹ A beta version of the dictionary is available at <https://leaddico.uclouvain.be>

(Granger et al. 2009), to identify learners' difficulties. As a large number of the latter are L1-specific, users are asked to specify their native language on the home screen so as to ensure that they are only presented with warnings and error notes that reflect their own potential difficulties. The LEAD is also a learning tool: it contains exercises and these too are customisable according to the user's L1. In addition, L1 customisation allows users to search in the dictionary via the translation of the English academic word in their mother tongue², a useful feature when users have difficulty recalling the English word. The dictionary also offers a second type of customisation: the examples of collocations and lexical bundles are automatically extracted from corpora representing the discipline that users have selected on the home page. The rationale for discipline customisation is that it is more useful and more motivating for learners to be presented with examples that match their writing situation as closely as possible. In addition, users have direct access to domain-specific corpora which they can use to search for any word, whether it features as an entry in the dictionary or not.

Error notes based on learners' authentic difficulties and tailored to their specific needs represent a significant advance on the intuition-based warnings that purportedly apply to all learners whatever their L1 background. However, they suffer from one major weakness that seriously diminishes their potential usefulness: learners will only look up words whose use they are unsure about, unaware that they may also need guidance on others. To solve this problem, a new functionality, called "highlighter", has been added to the LEAD. This dynamic interface not only hyperlinks all the academic words used in the user's text to their corresponding entries in the dictionary, but also, and more importantly, draws users' attention to potential (highlighted in orange) or real (highlighted in red) errors and displays explanations in the form of pop-ups. This feature of the dictionary is very much work in progress and is only implemented for French L1 learners at this stage.

Monolingual or bilingualised dictionaries are not the only types of dictionary that can benefit from learner corpus data. Corpus-based error notes or warnings are practically non-existent in bilingual dictionaries and yet this is arguably where they would be the most useful and easiest to implement as, unlike monolingual dictionaries, bilingual dictionaries are by their nature non-generic. However, including them would require a departure from the traditional bilingual dictionary model which caters for both L1 and L2 users of the two languages and serves both decoding and encoding purposes. A more flexible model consists in customising the dictionary to the L1 vs L2 status of the user and their encoding vs decoding purpose, as suggested by Thompson (1987: 284-5): "The dictionary should, like monolingual dictionaries, be aimed in one direction (e.g. for Chinese learners of English) rather than, like most bilingual dictionaries, in two directions (e.g. for Chinese learners of English and English learners of Chinese)." The idea of a "learner's bilingual dictionary" (Granger & Lefer 2016) is fully in line with this model. A unidirectional L1 → L2 encoding dictionary could include a host of useful error and usage notes reflecting learners' attested translation difficulties. These could, for example, include warnings against false friends and calques and, where appropriate, highlight the prevalence of translations by means of different grammatical categories and zero translations. The recently launched *Multilingual Student Translation* (MUST) collaborative initiative (Granger & Lefer 2018), which aims to collect translations produced by students in a wide range of languages, will be a particularly rich source for these notes.

4 Conclusion

Corpora have brought about a genuine watershed in lexicography, but the field has not yet reaped the full benefit of what they have to offer. In bilingual and pedagogical lexicography in particular, there is

² Currently Chinese, Dutch, French, German and Spanish

scope for more ambitious corpus-based and corpus-driven approaches. But there are nevertheless reasons to be optimistic about the future. New corpus collection initiatives will contribute to remedying the corpus shortage that the field is currently facing. For example, the *International Comparable Corpus* (Kirk & Čermáková 2017) will provide fully comparable corpora in nine languages. The recently released *Spoken BNC2014* (McEnery et al. 2017), which contains 11.5 million words of transcribed informal British English conversation, will provide much-needed up-to-date spoken data for English. There is unfortunately no collaborative initiative aimed at collecting large, representative translation corpora. This would, however, be entirely feasible as there are so many translated books available in electronic format. If several publishers were willing to lead this initiative in collaboration with academic research teams under the aegis of reputed associations such as EURALEX, ASIALEX, AFRILEX and the DSNA, bilingual lexicography could make a quantum leap forward. As to learner corpora, the *Multilingual Student Translation* corpus (Granger & Lefer 2018) and its standardised annotation system will be a very rich source of information on students' translation difficulties in a wide range of languages. Even now, however, use of available corpus resources and techniques, even very simple ones such as n-gram extraction, can already lead to substantial improvements. In particular, learner corpora can play a significant role in the new trend towards customisation of dictionaries given their ability to help identify language difficulties that are typical of particular learner groups. These insights were difficult to integrate into paper dictionaries but can be fitted seamlessly into web-based tools.

On a final note, it is salutary to remember that corpora are just collections of language data – undoubtedly data with high potential value, but data nonetheless. Even pre-analysed by powerful software tools, their use in lexicography will always require human expertise. As Moon (2008: 334) wisely reminds us, “lexicographers still have to use intuition and judgement in selecting, interpreting, and setting out the evidence, rather than simply relaying it to the user as quasi-scientific truth”.

References

- Adamska-Sałaciak, A. (2010). Examining equivalence. In *International Journal of Lexicography* 23(4): 387-409.
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow, Pearson Education.
- Bowker, L. (2010). The contribution of corpus linguistics to the development of specialized dictionaries for learners. In Fuertes-Olivera, P. (ed.) *Specialized Dictionaries for Learners*. Berlin & New York, pp. 155-168.
- Farina, A. (2009). Problèmes de traitement des “pragmatèmes” dans les dictionnaires bilingues. In: Heinz, M. (ed.). *Le dictionnaire maître de langue*, Berlin: Frank & Timme, pp. 245-264.
- García-Salido, M., Garcia, M., Villayandre, M., Alonso-Ramos, M. (2018). A lexical tool for academic writing in Spanish based on expert and novice corpora. *LREC 2018: Eleventh International Conference on Language Resources and Evaluation*, May 7-12, 2018, Miyazaki, Japan, pp. 260-265.
- Gouws, R.H. (2010). The presentation and treatment of collocations as secondary guiding elements in dictionaries. *Lexikos* 25, pp. 170-190.
- Granger, S. (1994). The learner corpus: a revolution in applied linguistics. In *English Today*, 39(10/3), pp. 25-29.
- Granger, S. (in press). Phraséologie et lexicographie bilingue: Apports croisés des corpus monolingues et parallèles. In Hanote, S. & Raluca, N. (eds.) *Autour de l'énonciation, de la lexicologie, de la morphophonologie et de la contrastivité : langues, discours, textes et corpus*. Presses universitaires de Rennes.
- Granger, S., Dagneaux, E., Meunier, F., Paquot, M. (2009). *The International Corpus of Learner English. Handbook and CD-ROM. Version 2*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Granger, S., Lefer, M.-A. (2012). Towards more and better phrasal entries in bilingual dictionaries. In *Proceedings of EURALEX 2012, August 2012*. Oslo, Norway, pp.682-692.
- Granger, S., Lefer, M.-A. (2013). Enriching the phraseological coverage of high-frequency adverbs in English-French bilingual dictionaries. In Aijmer, K., B. Altenberg, B. (eds.) *Advances in corpus-based contrastive linguistics*, Amsterdam & Philadelphia: Benjamins, pp. 157-176.

- Granger, S., Lefer, M.-A. (2016). From general to learners' bilingual dictionaries: Towards a more effective fulfilment of advanced learners' phraseological needs. In *International Journal of Lexicography* 29(3), pp. 279-295.
- Granger, S., Lefer, M.-A. (2018). MUST: A collaborative corpus collection initiative for translation teaching and research. Paper to be presented at the Using Corpora in Contrastive and Translation Studies (UCCTS2018) Conference, 12-14 September, 2018, Louvain-la-Neuve: University of Louvain.
- Granger, S., Paquot, M. (2015). Electronic lexicography goes local: Design and structures of a needs-driven online academic writing aid. In *Lexicographica* 31(1), pp. 118-141.
- Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Jackson, H. (ed.) (2013). *The Bloomsbury Companion to Lexicography*. London & New York: Bloomsbury.
- Kirk, J., Čermáková, A. (2017). From ICE to ICC: The new International Comparable Corpus. In: Bański, P., Kupietz, M., Lungen, H., Rayson, P., Biber, H., Breiteneder, E., Clematide, S., Mariani, J. Stevenson, M., Sick, T. (eds.) *Proceedings of the Workshop on Challenges in the Management of Large Corpora and Big Data and Natural Language Processing*. Birmingham, 24 July 2017. Mannheim: Institut für Deutsche Sprache, pp. 7-12.
- Maingay, S., Rundell, M. (1987). Anticipating learners' errors – implications for dictionary writers. In A.P. Cowie (ed.) *The Dictionary and the Language Learner*. Tübingen: Niemeyer, pp. 128-35.
- Matuška, O. (2018). Bilingual functionalities of Sketch Engine. In Granger, S., Lefer, M.-A. (eds.) *Book of Abstracts of the Fifth Using Corpora in Contrastive and Translation Studies Conference*, Louvain-la-Neuve, 12-14 September 2018. Louvain-la-Neuve: CECL Research Papers 1.
- McEnery, T., Love, R., Brezina, V. (2017). Compiling and analysing the Spoken British National Corpus 2014. Special issue of *International Journal of Corpus Linguistics* 22(3).
- Mogorrón Huerta, P. (2011). Compétences phraséologiques et traitement des expressions figées dans les dictionnaires. In: Van Campenhout, M., Lino, T., Costa, R. (eds) *Passeurs de mots, passeurs d'espoir : lexicologie, terminologie et traduction face au défi de la diversité*. Paris, Éditions des archives contemporaines, pp. 517-535.
- Moon, R. (2008). Dictionaries and collocation. In Granger, S., Meunier, F. (eds.) *Phraseology: An Interdisciplinary Perspective*. Amsterdam & Philadelphia: Benjamins, pp. 313-336.
- Nesi, H. (2014) Dictionary use by English language learners. *Language Teaching*, 47(1), pp. 38-55.
- Palmer, H. (1917). *The Scientific Study and Teaching of Languages*. University College, London. New York, World Book Company.
- Rundell, M. (1998). Recent trends in pedagogical lexicography. In *International Journal of Lexicography* 11(4), pp. 315-342.
- Rundell, M. (2007). The dictionary of the future. In Granger, S. (ed.) *Optimizing the Role of Language in Technology-Enhanced Learning*. Proceedings of the expert workshop organised by the Integrated Digital Language Learning seed grant project, Louvain-la-Neuve, 4-5 October 2007, pp. 49-51.
- Rundell, M. (2008). The corpus revolution revisited. In *English Today* 24(1), pp. 23-27.
- Rundell M., Granger, S. (2007). From corpora to confidence. *English Teaching Professional* 50, pp. 15-18.
- Rundell, M., Stock, P. (1992). The corpus revolution (Part 1). In *English Today* 8(2), pp. 9-14.
- Rundell, M., Stock, P. (1992). The corpus revolution (Part 2). In *English Today* 8(3), pp. 21-32
- Rundell, M., Stock, P. (1992). The corpus revolution (Part 3). In *English Today* 8(4), pp. 45-51.
- Salkie R. (2008). How can lexicographers use a translation corpus? In R. Xiao, L. He & M. Yue (eds) *Proceedings of The International Symposium on Using Corpora in Contrastive and Translation Studies*. Zhejiang University, Hangzhou, 25-27 September 2008.
- Sinclair, J. 2010 [2007]. Defining the Definiendum. In de Schryver, G.-M. (ed.) *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*. Kampala: Menha Publishers, 37-47.
- Tarp, S. (2009). Reflections on data access in lexicographic works. In S. Nielsen, S. Tarp (eds.) *Lexicography in the 21st Century: In Honour of Henning Bergenholtz*. Amsterdam & Philadelphia, Benjamins, pp. 43-62.
- Thompson, G. (1987). Using bilingual dictionaries. *English Language Teaching Journal* 41(4), pp. 282-286.
- Verlinde, S., Peeters, G. (2012). Data access revisited: The Interactive Language Toolbox. In: Granger, S., Paquot, M. (eds) *Electronic Lexicography*. Oxford: Oxford University Press, pp. 147-162.
- Xia, L. (2015). Corpora and collocations in Chinese-English dictionaries for Chinese users. *English Language Teaching* 8(10), pp.162-167.