



Issues in corpus-informed research and learning in ESP

Alex Boulton, Shirley Carter-Thomas, Elizabeth Rowley-Jolivet

► To cite this version:

Alex Boulton, Shirley Carter-Thomas, Elizabeth Rowley-Jolivet. Issues in corpus-informed research and learning in ESP. Alex Boulton; Shirley Carter-Thomas; Elizabeth Rowley-Jolivet. Corpus-informed research and learning in ESP : issues and applications, 52, John Benjamins, pp.1-14, 2012, Studies in corpus linguistics, 978-90-272-0357-1. 10.1075/scl.52.01bou . hal-00683783

HAL Id: hal-00683783

<https://hal.science/hal-00683783>

Submitted on 7 Aug 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Alex Boulton, Shirley Carter-Thomas & Elizabeth Rowley-Jolivet. 2012. Issues in corpus-informed research and learning in ESP. Introduction to A. Boulton, S. Carter-Thomas & E. Rowley-Jolivet (eds), *Corpus-Informed Research and Learning in ESP: Issues and Applications*. Amsterdam: John Benjamins, p. 1-14. DOI: 10.1075/scl.52.01bou

Alex Boulton, Shirley Carter-Thomas & Elizabeth Rowley-Jolivet. 2012. Issues in corpus-informed research and learning in ESP. Introduction to A. Boulton, S. Carter-Thomas & E. Rowley-Jolivet (eds), *Corpus-Informed Research and Learning in ESP: Issues and Applications*. Amsterdam: John Benjamins, p. 1-14. DOI: 10.1075/scl.52.01bou

1. Introduction

The vast majority of students currently studying English in European higher education are enrolled in disciplines other than English such as science, engineering, medicine and economics, or in cross-disciplinary areas such as specialised translation and terminology. The relevance of tertiary courses in English for Specific and/or Academic Purposes (ESP/EAP) to the professional needs of these students – mastering the genres and appropriate specialised registers of English – can be greatly enhanced by a corpus-informed approach. Corpus analysis provides an empirically-based understanding of the discourse and language used for specific purposes in the students' own fields, providing them with greater autonomy for carrying out their professional tasks.

To date, however, while corpora have been widely used for research purposes in general English (lexicography, phraseology, etc.) and in specialised language and genres, little of this has yet had any concrete impact on the material and pedagogy of ESP courses or on the students' involvement in the learning process. ESP is no different in this respect from English as a Foreign Language (EFL) in general, as a similar lack of cross-fertilisation between corpus research and teaching applications in EFL or Second Language Acquisition (SLA) is a recurrent leitmotiv in many recent studies (e.g. Mukherjee 2004; Römer 2009). As Aijmer (2009: 2) succinctly puts it, “the direct exploitation of corpora in the EFL classroom is unusual and the impact of corpora on syllabus and materials design has been slight.”

At first sight, the picture appears more encouraging at the graduate and postgraduate levels. With English now the sole language of research publication in many academic disciplines, a large number of studies, many of them corpus-based, have been carried out in recent years into various research genres to provide support for academics who have to publish in English, not their native language. While this has increased our understanding of disciplinary discourses immeasurably, here also the application of corpora to university ESP courses is a rarity, and the independent exploitation of specialised language corpora by non-linguist academics themselves (scientists, economists, etc.) appears to be sporadic. Despite the often positive feedback from participants from courses in corpus compilation and consultation in their disciplinary speciality, the longer-term take-up of these techniques has yet to be confirmed (cf. Lee and Swales 2006; Chambers 2007: 13).

This is all the more regrettable in that corpora and corpus-informed learning are arguably even more useful and relevant in the ESP context – and more widely, in Languages for Specific Purposes (LSP) – than in EFL, given the characteristic features of ESP itself. The first is the typical ESP learner profile: corpus consultation calls on more learner-centred, individualised and autonomous methods of learning, such as hypothesis building and an inductive, problem-solving approach suitable for learners who are already used to research procedures in their own field (Johns 1986: 161). There is evidence (Johansson 2009; Kennedy & Miceli 2001) that this type of method is more successful with adult learners – the case in

ESP, where most learners and users are in higher education or beyond. The second important feature concerns the lexico-grammar of ESP language varieties. While they obviously conform to the syntax and semantics of English as a whole, these varieties can be distinguished from general English by their selective use of certain structures, the prevalence of domain-specific, often highly conventionalized, phraseologies (collocations, lexical bundles), and the extremely rapid evolution of ESP terminology and lexis to keep pace with technical and professional developments. For serious study of all these features, corpora are indispensable: information about the frequency of use of certain structures, and about specialised phraseologies and patterns, can only be obtained from corpora, not from textbooks or grammar books, while traditional dictionaries cannot compete with web-based corpora where lexical and terminological evolution is concerned.

This type of information is particularly important for second- or foreign-language (L2) ESP users, at whatever level. They have for the most part received training in general English, which for the reasons outlined above proves far from sufficient to meet their professional or domain-specific language needs. A further problem concerns their awareness of register and genre differences: knowing which lexico-grammatical forms and discourse conventions are appropriate in a given context (writing *vs* speech, academic and other professional genres *vs* more personal forms of expression, etc.) is highly problematic for L2 users even at advanced levels, and lack of awareness in this area can reflect negatively on their professional competence.

The relevance of corpora to ESP does not stop here. Ever since the inception of ESP as a research field in its own right, which Swales in *Episodes in ESP* (1985) traces back to Charles Barber's 1962 paper "Some measurable characteristics of modern scientific prose", the contrastive approach has been a dominant paradigm. By comparing usage in specialised and general language, between disciplines or fields of activity, or between different time-periods and between different levels of reader / writer expertise, many of the specificities of professional discourses and epistemologies have been detected, while cross-linguistic studies have revealed culture- and language-specific differences between LSPs. Corpora are ideally suited to contrastive analysis, and therefore particularly applicable in ESP, a methodological advantage that applies equally well to quantitative and to qualitative studies. For quantitative investigations, sophisticated statistical packages (several of which are used in the chapters in this collection) are now available for corpus applications, making fine-grained analysis possible for large volumes of data or contrasted subsets in the corpus. For more qualitative approaches, contrastive corpora are very useful for establishing the salient features of particular language varieties.

For practical reasons also, corpora are well-suited to ESP, both for research and for learning applications. As several studies have pointed out (e.g. Aston 1997; Ghadessy et al. 2001; Gavioli 2002; Flowerdew 2008), small can be beautiful, i.e. significant results can be obtained using small corpora of various kinds, provided they are well constructed. The more specialised and specific the domain covered, the smaller the corpus can be. This is also a decided advantage for learners of ESP, where needs are often highly specific and there are many relatively small professional and disciplinary communities. Corpora can be tailored to the learners' specific requirements with smaller-scale disciplinary corpora compiled for

targeted research projects. These small corpora can then use the vast language banks compiled for lexicography and analysis of general language as reference corpora in order to highlight differences in specialised usage.

A criticism often levelled at corpus analysis is that while it is an extremely powerful approach for ‘low-level’ phenomena such as lexical and grammatical items, it is ill-adapted to ‘higher-level’ considerations of discourse organisation, argumentation, and genre, or to a text linguistics approach, all of which are of prime importance in ESP. As Flowerdew (2002) demonstrates, however, corpus linguistics is now moving away from its early lexicogrammatical focus towards a broader perspective encompassing discourse phenomena, either by combining bottom-up and top-down approaches (Charles 2007; Flowerdew 2008) or by developing special tools to investigate discourse patterns (Biber et al. 2007). The field of corpus-based ESP analysis is thus evolving in promising directions and being gradually enriched by new methods and applications, and, true to its origins, by empirical investigations that have a robust theoretical foundation.

2. Overview of the volume

All these considerations are addressed in the present volume, which comprises specially-commissioned studies covering a range of corpus-informed approaches to researching, teaching and learning ESP. The chapters cover a wide variety of disciplines, ranging from medicine, biology, earth sciences and physics, to economics and linguistics, and a range of ESP users: specialised translators, undergraduates, junior and experienced researchers, and language trainers. The types of corpora compiled or used vary in size from small tailor-made collections of a few thousand words to published corpora amounting to hundreds of millions of words. They are also highly varied in nature, comprising contrastive corpora of various kinds: English vs French, native vs non-native speaker, general English vs ESP, mono- vs interdisciplinary, written vs spoken, review vs research articles, and corpora for “colony texts” (Hoey 2001) such as specialised dictionaries.

The different contributions also exemplify an interesting and often novel range of methodological and theoretical approaches to ESP corpus analysis: semantic prosody, text statistics and lexicometry, standard deviation, move analysis, phraseology, systemic functional linguistics, qualitative and quantitative approaches, and prototype theory. The levels and scope of the analyses presented likewise cover the spectrum from individual linguistic items (personal pronouns, adjectival and nominal modifiers), to communicative metafunctions and rhetorical moves, finishing with a meta-analysis of ESP corpus research in language learning. The breadth of domains and disciplines, corpus types and methodologies enables discussion of a considerable number of issues prevalent in ESP and corpus research, including those of corpus size and representativity, contextual relevance, and the role of ESP teachers with regard to indirect applications of corpus findings and direct uses of corpora for language teaching and learning.

The common point which counterbalances this variety and unites all the chapters in the volume is the focus on ESP users from a single mother tongue (L1) background (French) in a single teaching / learning context. The contributors are all associated with GERAS (*Groupe d'Etude et de Recherche en Anglais de Spécialité*, the French national research group in ESP),

and the majority are involved in higher education in France. This shared background gives a deeper insight into the interrelationships between L1 and L2, culture, learning environments and methods, as the problems encountered by this representative ESP user group are considered from different angles, providing a richer, convergent picture rather than isolated snapshots of heterogeneous language groups and cultures. The scope of the volume is not restricted to France, however, , as the methodologies proposed are not limited to the French context and will be relevant to teachers, researchers and course designers in many different countries looking at ESP from different language perspectives.

Whilst an interest in applied corpus use underlies all the chapters, they have been grouped into three broad sections depending on their dominant focus: language, genre, or pedagogy. These are presented in more detail below.

3. ESP corpora for language research

The first section explores how corpus analysis can enhance descriptions of specialised English. The contributions highlight certain specificities of language use within carefully defined ESP contexts, and suggest ways in which both the results and the methods can be further exploited by course designers, translators, teachers and students.

The first chapter sets the scene for a comparative approach, looking at the use of first person pronouns in comparable corpora of introductions to economics research articles (RAs) in French and English from the KIAP corpus. **Shirley Carter-Thomas** and **Angela Chambers** show how corpus linguistics is not limited just to the traditional questions of lexico-grammar, but can be equally useful in discourse analysis. They begin with a qualitative analysis of one RA introduction in each language, relating pronoun use to Swales' (1990) CARS model as writers in both languages first establish a territory, then establish a niche, and finally occupy that niche. The following quantitative analysis of co-occurring verbs also reveals different uses of the pronouns, confirmed in the concordances: most frequently to represent author as researcher, followed by author as writer (use of discourse verbs), and occasionally author as arguer (position verbs). The similarities in use between the two languages are striking, though it is notable that English uses the first person singular far more frequently than French in this genre, thus contradicting the conventional wisdom that is passed on to novice writers of English. The authors argue that teachers and learners can, with comparative ease, create and consult relevant corpora in their own disciplines, combining qualitative and quantitative techniques to discover such uses for themselves: the advantage lies not just in the results obtained, but in the process of uncovering them and thus constructing their own discipline-specific knowledge.

Another investigation into how corpus linguistics can help with discourse is provided by **Anthony Saber** in Chapter 2, which focuses on the somewhat larger linguistic units of phraseological patterns. Ever since Sinclair's (1991) seminal insight that a large part of our language activity is based on the idiom principle (as opposed to the open-choice principle of constructing sentences word by word), it has become increasingly clear that specialized discourses in particular make very intensive use of recurrent collocations and stereotypical phraseological units. Starting from this acknowledged feature, the originality of Saber's study lies in the connections he establishes between these phraseological patterns and the

rhetorical moves and steps found in research articles. Using a specially compiled million-word corpus of biomedical RAs, sub-divided into IMRAD sections (Introduction, Methods, Results, Discussion), Saber first identifies the key words in the corpus as a whole, and then the most salient items in each IMRAD section (i.e. those which occur significantly more frequently in one or more sections), which are subjected to a cluster analysis. This procedure shows how researchers fine-tune the phraseological patterns to each IMRAD section and each rhetorical step: not only does each section have its own relatively predictable set of moves and functions, but the lexical and phraseological choices are remarkably consistent within sections, a finding which would go largely unnoticed by looking only at the corpus as a whole. A further interest of the study is that it draws both on rhetorical discourse analysis and on the RA structure recommended by the international methodological guidelines for medical research – the CONSORT statement for randomized controlled trials and the STROBE statement for epidemiological studies – which an increasing number of medical journals require their authors to comply with, thus firmly anchoring the analysis in professional practice.

Chapter 3 by **François Maniez** remains within the medical field but takes a contrastive perspective in order to address noun phrase modification. It has frequently been attested that nominal premodification (N+N, e.g. *liver function*) is, alongside adjectival modification (ADJ+N, e.g. *hepatic function*), a prominent feature of specialized discourse. This is a pitfall for non-native users of ESP, however, especially from Romance language backgrounds which do not allow premodification by a noun, using adjectival or prepositional modification instead. The choice between a premodifying noun or adjective is therefore difficult to make for non-native writers of English, particularly as general English often differs markedly in this respect from specialized usage, where issues of genre and register arise. By comparing the results obtained from a detailed analysis of some N+N and ADJ+N pairs in three corpora – the multilingual specialized medical corpus EMEA, the general and specialized sub-registers of CoCA, and the web-as-corpus – Maniez shows how corpus consultation can help solve this dilemma for ESP learners, and can provide a reliable guide in selecting the type of modification that is appropriate to disciplinary use. The approach can be applied to ESP fields other than medicine and to languages other than French, and it is argued that non-native ESP writers and specialized translators can also greatly benefit from initiation into corpora such as CoCA and into the advanced query syntax of web search engines.

Specialised translation is also the focus of the final chapter in this section by **Natalie Kübler** and **Alexandra Volanschi**, who tackle the issue of semantic prosody. Corpus linguistics, via analysis of collocates, has highlighted how individual words can have positive or negative semantic prosody by associating with strongly connotated terms, projecting their connotations and evaluative meaning onto the surrounding context. Semantic prosody is difficult, however, for non-native speakers of a language to perceive. When translating from a second language into their mother tongue, not recognising semantic prosodies may lead to misunderstanding the intent of the source text, and consequently mistranslating it in the target language. Despite its relevance to translation studies, investigations of this phenomenon have to date been mainly restricted to English, and have concerned general English rather than ESP. Kübler and Volanschi here propose a cross-linguistic analysis of the semantic prosody of cognate words in English and French texts in earth sciences in order to

test the pervasiveness of semantic prosody, i.e. whether it applies with equal force to LSPs as to non-specialised language. The results of their analysis of verb pairs such as *to commit / commettre*, or *to cause / causer*, shows that the negative semantic prosody of these items in general usage in both languages does not necessarily carry over to specialised discourse, where it is smoothed. In the English corpus, most noun collocates are neutral, especially where they are field-specific technical terms; where negative prosody is conserved, this is usually because it has an effect on human beings. However, intuitions of general semantic prosody are likely to influence use in specialised contexts for both native and non-native speakers or translators, and may result for example in the unnecessary avoidance of terms in ESP simply because they are strongly connoted in general language. The paper thus argues for the use of comparable (as opposed to parallel translation) corpora, and the inclusion of information on semantic prosody in terminological databases for translators.

4. ESP corpora for genre-based approaches

One of the most fruitful approaches to ESP in the last two decades, both for research and teaching purposes, has been genre analysis, which investigates the typified and recurrent features of specialized language and discourse. Corpus-based methods are particularly relevant and useful in genre analysis as they enable these features to be reliably identified. It is not surprising, therefore, that genre considerations are a leitmotiv throughout this volume; in the second section, however, they become the main focus of attention.

Unlike the other studies in the volume which use written corpora, the chapter by **Elizabeth Rowley-Jolivet** investigates an oral genre: the research conference presentation in science. Using a purpose-built corpus, the author contrasts the highly condensed expression on the slides with the much more extended oral commentaries from a systemic-functional perspective. She first examines grammatical metaphor (in particular, the nominalisation of verbal processes), generally considered to be typical of scientific writing but not of speech, and finds that while indeed the slides are characterised by a heavy use of grammatical metaphor, a simple written / spoken contrast does not hold: this phenomenon is also a significant feature of the commentaries in this research genre, at least for material processes, for reasons related to topic, audience, and epistemology. She next considers the three communicative metafunctions – ideational, textual, and interpersonal – in the two subsets. This analysis clearly reveals that each is functionally specialised: the ideational metafunction dominates on the slides, which convey almost exclusively the informational content of the talk, whereas the spoken commentaries fulfil both the textual and interpersonal metafunctions by creating cohesion, evaluating the data, and making audience contact. In sum, the slides give the disciplinary substance of the talk, the commentaries weave it all together into coherent discourse. A key concept in the systemic-functional approach is the question of choice: this corpus study shows that conference speakers are faced with options between alternative wordings when oralising their research, requiring on their part a high degree of lexico-grammatical flexibility to adapt to the constraints of the two modes of presentation. Novice researchers (especially non-native speakers) need to be aware of the specificities of each, which will not be covered if they are taught only how to read and write research articles in their field.

Genre specificities can be approached from several perspectives. In the following chapter **Céline Poudat** and **Peter Follette** concentrate on quantitative and text-statistics methods to highlight disciplinary variations within the research article genre. Using a corpus of research articles tagged by part of speech, they first examine variation between the academic fields of biology and linguistics, two disciplines that differ widely with respect to their experimental approaches, methodology, and intellectual communities and history. Focusing on the highly specific lemmas detected by the software, they show that experimental discourse such as biology is characterized by content markers that reflect specialized terminology, whereas speculative discourse such as linguistics uses a much wider variety of function words and patterns that signal a particular style. In the second stage, using the same methodology of hypergeometric distributions and high specificities, the authors analyse genre variation by comparing review articles and research articles within their biology corpus. Although the differences here are less marked than in the domain comparison, the results show clear associations between each genre and parameters such as verb tense, general vs specific scientific terms and use of *we*. Correspondence factor analysis, which assesses distances between observations, is then applied to the whole corpus and reveals that the different subsets fall into highly distinct zones of the factor map. These quantitative approaches are followed up by a more qualitative focus on markers of rhetorical moves which confirms the statistical results: the different functions of an article, and the lexical elements used to achieve them, differ significantly between different academic fields and between similar but distinct genres. This leads the authors to highlight in their conclusion the difficulties for teachers expected to teach ‘general’ academic English, and to recommend a greater recourse to ESP corpora.

A methodology for assessing genre variation is also the topic of **Dacia Dressen-Hammouda**’s study of field accounts, a part-genre found in research articles in geology. Using a diachronic corpus composed of the publications of five professional geologists over a ten-year period following their PhD, she examines how scientific writers’ disciplinary voice shifts over time as a result of increasing expertise, as they progress from being junior to senior researchers. In order to measure these shifts in voice, she proposes a method to date unused in genre-based corpus analysis: a measure of standard deviation, which compares changes in the authors’ individual writing strategies over a period of time with norms identified in the corpus. Dressen-Hammouda argues that genre use is characterized in fact by substantial variation from perceived norms, since a norm does not represent a single value *per se*, but an entire range of values accepted by the disciplinary community in question, and that these deviations from the norm can be tracked and quantified using standard deviation. The corpus is analysed for 13 variables grouped into three categories of indicators: ‘personalisation cues’ which signal author agency, ‘doing-the-work cues’ which allow the reader to infer that fieldwork has been carried out by the author, and ‘disciplinary cues’ which demonstrate the author’s mastery of the epistemological framework of the discipline. The results show that while the overall distribution of the variables remains fairly stable as the authors progress in their discoursal expertise, there is a general tendency to make increasing use of personalisation cues and to bend the norms as writers construct their credibility and authority and develop their own disciplinary voice. It is particularly notable that most of the writers do indeed develop along very similar trajectories. The study raises a

number of interesting questions, both theoretical and applied, for corpus-based genre analysis, and argues that a genre corpus reflects the ‘typicality’ of a particular genre, based on the most frequently used strategies, but does not reflect individual differences in disciplinary expertise and voice.

5. ESP corpora for language teaching and learning

One of the most obvious direct applications of corpus linguistic description is to language learning and teaching, especially in LSP, and this is the topic of the final section of the volume. First, corpora can be of tremendous use in reference resources such as textbooks, grammars, usage manuals and dictionaries. In such cases, their input may be all but invisible to teacher and learner as end-users (cf. McCarthy 2004), but increasingly technology allows easy interaction with overtly corpus-derived data. Second, they can be used as (the source of) materials for teachers to work on and learners to learn from. These uses may be more explicit, though they are not necessarily restricted to the traditional concordances and wordlists, encompassing also text-based work for enhanced discourse awareness. Whatever the approach – ‘indirect’ or ‘direct’ corpus-based work (e.g. Römer 2010) – it is not enough for teachers to adopt the ‘armchair’ approach and simply assume success based on intuition, experience, and anecdote. Rather, methodical investigation is required to examine the alleged benefits and weigh them against the attested difficulties. These are the topics covered in the three chapters in this section.

Corpora can inform materials, but even corpus-based dictionaries and word lists are open to criticism on a number of counts, a point discussed in the first chapter in this section by **Geoffrey Williams**. Amongst other things, advanced learner’s dictionaries mainly help with reception, not production, and focus largely on English for general rather than specific purposes, the problem being exacerbated in bilingual dictionaries with even higher levels of decontextualisation. Coxhead’s (2000) Academic Word List has also recently come in for some criticism for being based on outdated sources (cf. Cobb 2010), and indeed any list (notably the Academic Formulas List: Simpson-Vlach & Ellis 2010; the Academic Collocation List: Ackermann et al. 2011) encounters problems due to its static nature. The research in this chapter shows how corpora can do more than simply inform traditional dictionary entries or other itemised lists: they allow collocational networks to arise ‘organically’ from the data rather than being imposed by the lexicographer. Context and use are thus paramount along the lines of the pattern dictionary (e.g. Hanks 2008), and the electronic format allows the user to access far more ‘examples’ (*pace* Gavioli 2005: 7) than traditionally available. The prototype approach also means that entries can be ordered by (partially) overlapping senses rather than alphabetically, providing the user with a much more accessible and interactive tool for productive purposes. The examples given here show what such an organic dictionary could look like on line, currently based on a 33 million-word medical corpus.

As the ‘publish or perish’ philosophy becomes increasingly commonplace around the world, many researchers experience pressure to write in English. Explicit training is the exception, and non-native speakers are expected to simply pick up the lexico-grammar and discourse of their discipline by themselves. In the first paper in this section, **Susan Birch-Bécaas** and **Ray Cooke** describe how a small multi-disciplinary corpus can help via their open-access on-line

tool TYOS (Type Your Own Script). The 150,000-word corpus includes case reports, letters of request and complaint, replies to reviewers and other written genres in addition to the research articles focused on here. The specificity of the corpus is that it includes first drafts of all texts in addition to the final versions, thus highlighting the writing process through the various modifications, with special functions for discourse moves, useful expressions, linkers, vocabulary, grammar, and so on. The domain-specific language is presented as 'bricks', with general academic language as 'cement' covering several disciplines, though the focus is on consciousness-raising rather than teaching. The tool itself can be used autonomously or, as here, in class, where the participants can work collaboratively to highlight the specificities of their own discipline, switching between whole texts and concordances, as well as Internet searches. Although both producers and consumers of the corpus here are French, it is likely that many of the features will be relevant to researchers from other L1 backgrounds. The work ties in well with other uses of corpora for ESP writing, such as Lee and Swales (2006), Hafner and Candlin (2007), or Sun (2007), as outlined in the next chapter.

The final chapter by **Alex Boulton** surveys empirical research evaluating corpus use for ESP learning – 20 studies out of a current total of more than 80 in 'data-driven learning' as a whole, to use the term coined by Johns (1990). One key feature to emerge is the remarkable variety in all areas – the corpora and tools used, the language objectives, the research questions and instruments, the learners and disciplines covered, and so on. Most objectives relate to writing (including revision, error-correction and translation) or vocabulary and lexico-grammar, with relatively few focusing on discourse and only one on speaking – a distribution reflected partly in the present book. In all of these areas and with virtually no exceptions, each individual study is promising, even if the results are not always statistically significant (to be expected in current dynamic systems theory; cf. de Bot 2008: 173). Though it is argued that the disparate nature of the various studies makes a formal meta-analysis all but impossible, the overall picture which emerges is extremely positive on the whole – not just in terms of learning outcomes, but in using corpora as a reference resource, in learners' attitudes and receptivity, and in what learners actually do with language corpora. The weight of evidence, it is argued, shows that corpora can be useful for many learners for many purposes at least some of the time, a powerful addition to the ESP teacher's and learner's repertoire when used sensitively and appropriately.

6. Perspectives

As this volume shows, corpus analysis techniques can be used profitably by a variety of language professionals as well as learners, to investigate areas in ESP of a more or less specialised nature, in response to their objectives or needs. In many of the chapters the traditional division between corpus description and corpus applications has tended to converge towards an approach combining research, learning and teaching. Quantitative and qualitative methods are also seen to usefully complement each other, with concordance lines or various statistical measures being fruitfully combined with context- and genre-sensitive approaches. The same techniques are applied to longitudinal studies of individual writers, to the different versions of a written document, or to the design of organic dictionaries. Language, including ESP, is constantly evolving and corpus techniques can help track these advances, avoiding reifying norms. Such research themes represent promising new departures.

Another common thread throughout the book is the focus on French ESP users and on the various cross-cultural and cross-linguistic issues this implies. The insight into L1 and L2 similarities and differences is something that should, we feel, remain prominent in ESP research. In all types of ESP – not just in research communication – it is essential to take into account students' own linguistic or cultural background when familiarising them with various ESP domains.

Looking farther ahead, what might the future hold for ESP? Corpus-based investigations such as the ones presented here are a clear indication of the impact that digital media and resources are having on language research and teaching. These media are already leading to the emergence of new genres and communicative practices in the professional domains studied in ESP, and represent a stimulating challenge for the field as we adapt our present methods and concepts of research and teaching to address these new forms of discourse and language. The humanities in general have yet to exploit the affordances of digital media to the full in their own practices, however, unlike some other disciplines studied here. Science, for instance, now makes intensive use of web-enabled linked data, through vast online repositories of information and open access to publications, enabling the reuse, repurposing, and rapid exchange of data and findings for new developments and increased collaboration. Digital humanities, in contrast, is still but a fledgling movement, in Europe at least. Although among the humanities, ESP has been quick to see the potential of some of these techniques and affordances, there is still a long way to go. Publication is still predominantly gated, and the primary data used in studies are generally hidden and inaccessible. This is clearly a big problem for corpus work and other language research, but also results in substantial duplication of research and teaching effort which slows down advances in the field, and, as the final chapter in this collection shows, in difficulties comparing findings from different studies. Our hope for the future of ESO is that digital media will create a greater drive towards more open access, more data-sharing of teaching and learning material and of the many interesting, often small, corpora used in the field, so as to encourage collaborative practice and move the discipline forward.

References

- Ackermann, K., Biber, D. & Gray, B. 2011. An academic collocation list. *Corpus Linguistics* 2011. Birmingham: Centre for Corpus Linguistics, 20-22 July.
- Aijmer, K. (ed.). 2009. *Corpora and Language Teaching*. Amsterdam: John Benjamins.
- Aston, G. 1997. Small and large corpora in language learning. In B. Lewandowska-Tomaszczyk & J. Melia. (eds) *Practical Applications in Language Corpora*. Lodz: Lodz University Press, p. 51-62.
- Barber, C.L. 1962/1985. Some measurable characteristics of modern scientific prose. In *Contributions to English Syntax and Phonology*, 1-23. Stockholm: Almqvist & Wiksell. Reproduced in *Episodes in ESP*, J. Swales (ed.), 1-16. New York: Prentice Hall.
- Biber, D., Connor, U. & Upton, T.A. 2007. *Discourse on the move: Using corpus analysis to describe discourse structure*. Amsterdam: John Benjamins.
- Chambers, A. 2007. Popularising corpus consultation by language learners and teachers. In *Corpora in the Foreign Language Classroom*, E. Hidalgo, L. Quereda & J. Santana (eds), 3-16, Amsterdam: Rodopi.

- Charles, M. 2007. Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes* 6(4): 289-302.
- Cobb, T. 2010. Learning about language and learners from computer programs. *Reading in a Foreign Language* 22(1): 181-200.
- Coxhead, A. 2000. A new academic word list. *TESOL Quarterly* 34(2): 213-238.
- de Bot, K. 2008. Second language development as a dynamic process: Introduction to the special issue on dynamic systems theory. *Modern Language Journal* 92(2): 176-188.
- Flowerdew, L. 2002. Corpus-based analyses in EAP. In J. Flowerdew (ed.), *Academic discourse*. London: Longman, 95-114.
- Flowerdew, L. 2008. *Corpus-Based Analyses of the Problem – Solution Pattern: A Phraseological Approach*. Amsterdam: John Benjamins.
- Gavioli, L. 2002. Some thoughts on the problem of representing ESP through small corpora. In B. Kettemann & G. Marko. (eds) *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi, 293-303.
- Gavioli, L. 2005. *Exploring Corpora for ESP Learning*. Amsterdam: John Benjamins.
- Ghadessy, M., Henry, A. & Roseberry, R. (eds.) 2001. *Small Corpus Studies and ELT: Theory and practice*. Amsterdam: John Benjamins.
- Hafner, C. & Candlin, C. 2007. Corpus tools as an affordance to learning in professional legal education. *Journal of English for Academic Purposes* 6(4): 303-318.
- Hanks, P. 2008. Why the ‘word sense disambiguation problem’ can’t be solved, and what should be done instead. In *Corpus Linguistics, Computer Tools, and Applications: State of the art*, B. Lewandowska-Tomaszczyk (ed.), 97-116. Frankfurt: Peter Lang.
- Hoey, M. 2001. *Textual interaction*. London: Routledge.
- Johansson, S. 2009. Some thoughts on corpora and second-language acquisition. In *Corpora and Language Teaching*, K. Aijmer (ed.), 33-44. Amsterdam: John Benjamins.
- Johns, T. 1986. Micro-Concord: A language learner’s research tool. *System* 14(2): 151-162.
- Johns, T. 1990. From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria* 10: 14-34.
- Kennedy, C. & Miceli, T. 2001. An evaluation of intermediate students’ approaches to corpus investigation. *Language Learning & Technology* 5(3): 77-90.
- Lee, D. & Swales, J. 2006. A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes* 25: 56-75.
- McCarthy, M. 2004. *Touchstone: From corpus to coursebook*. Cambridge: Cambridge University Press.
<http://www.cambridge.org/us/esl/Touchstone/teacher/images/pdf/CorpusBookletfinal.pdf>, accessed 14/04/08.
- Mukherjee, J. 2004. Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In *Applied Corpus Linguistics: A multidimensional perspective*, U. Connor & T. Upton (eds), 239-250. Amsterdam: Rodopi.
- Römer, U. 2009. Corpus research and practice: What help do teachers need and what can we offer? In *Corpora and Language Teaching*, K. Aijmer (ed.), 83-98. Amsterdam: John Benjamins.

Alex Boulton, Shirley Carter-Thomas & Elizabeth Rowley-Jolivet. 2012. Issues in corpus-informed research and learning in ESP. Introduction to A. Boulton, S. Carter-Thomas & E. Rowley-Jolivet (eds), *Corpus-Informed Research and Learning in ESP: Issues and Applications*. Amsterdam: John Benjamins, p. 1-14. DOI: 10.1075/scl.52.01bou

- Römer, U. 2010. Using general and specialised corpora in English language teaching: Past, present and future. In *Corpus-Based Approaches to English Language Teaching*, M.-C. Campoy, B. Belles-Fortuno & M.-L. Gea-Valor (eds), 18-35. London: Continuum.
- Simpson-Vlach, R. & Ellis, N.C. 2010. An academic formulas list: New methods in phraseology research. *Applied Linguistics* 31(4): 487-512.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sun, Y.-C. 2007. Learner perceptions of a concordancing tool for academic writing. *Computer Assisted Language Learning* 20(4): 323-343.
- Swales, J.M. 1985. *Episodes in ESP*. New York: Prentice Hall.
- Swales, J.M. 1990. *Genre Analysis*. Cambridge: Cambridge University Press.