Chapter 1

Introduction to Corpus Linguistics and ELT

Mari Carmen Campoy-Cubillo, Begoña Bellés-Fortuño and Maria-Lluïsa Gea-Valor

Universitat Jaume I, Castelló

From its origins, Corpus Linguistics has had a strong link with language teaching. John Sinclair’s impact on dictionary making and his pioneering work on corpus research (Sinclair 1987, 1991, 2004) have been the starting point for many corpus-based approaches to language teaching (Wichmann et al. 1997; Burnard and McEnery 2000; Granger et al. 2002; Kettemann and Marko 2002; Aston et al. 2004; O’Keefe et al. 2007; Aijmer 2009, to name but a few). The common ground for all these approaches is that they are based on empirical evidence, thus leading to the elaboration of better quality learner input and providing teachers and researchers with a wider, finer perspective into language in use, that is, into the understanding of how language works in specific contexts.

Corpus-Based Approaches to ELT presents work by leading linguists exploring different ways of applying corpus-based and corpus-informed research to language teaching environments. More specifically, the volume tackles three main areas of special interest today: the use of corpora for teaching English for Specific Purposes, pedagogically motivated uses of corpora, and the potential of corpora-mediated multimodal tools for the language learning context.

The compilation, description and analysis of domain-specific corpora is one of the widest areas of research in corpus linguistics, especially as regards academic and professional settings. This book provides an in-depth analysis of academic and professional texts by means of corpus-based methodologies in order to enhance English for Specific Purposes (ESP) teaching. A wide perspective into ESP corpora is offered, as the chapters include written and spoken academic discourse, the use of English language in professional contexts, and the use of both native English speaker corpora and ESP learner corpora, that is, corpora in which learners attempt at producing professional texts.
The second issue examined in this volume has to do with how English language teaching may benefit from corpus data to improve language learner input (the so-called corpus-based and corpus-informed approaches) and the different ways in which corpora may aid in understanding learner and teacher discourse. In this sense, the volume illustrates the way corpora may be used directly in the classroom and how corpus research may be applied to inform syllabi and classroom materials.

Finally, the third dimension reflects on the role of corpus tools and multimodal devices, where corpora-based research plays a central role to inform teaching materials. Multimodal corpora are still in their infancy when compared to corpora where only one discourse mode is used. Challenges in this area lie not only in the design of such corpora, a difficult task per se, but also in the reflection on how information is organized and connected among the different text modes. Far from being just an inclusion of one or more corpora within a learning package and allowing users access to concordance and collocational information, this entails having a clear idea of the pedagogical goals of both tool and tool applications and how corpora are integrated in the tasks a learner is intended to carry out. It also implies a lot of research into feasible text mode combinations and consensus on issues such as possible tagging categories and terminology in order to be able to contrast studies carried out by different researchers.

The volume opens with Ute Römer’s chapter, in which she presents and discusses the state of the art in the field of corpus linguistics and language teaching. The author provides an overview of the past, present and future developments in corpus linguistics, reviewing the applications of general and specialized corpora. Römer insightfully points at the need to foster the use of pedagogical corpora and draws a work agenda around three main topics: focus on learner and teacher needs, indirect uses of corpora in language teaching and direct uses of corpora in language teaching.

From this introductory chapter, the volume goes on to study the close relationship between corpus linguistics and language teaching, and is divided into three more Parts, namely Corpora and English for Specific Purposes; Learner Corpora and Corpus-Informed Teaching Materials; and Multimodality: Corpus Tools and Language Processing Technology.

1.1 Corpora and English for Specific Purposes

Part I of this book contains six chapters describing various scenarios related to the field of *English for Specific Purposes* (ESP), including academic and
professional settings. Although both learner and expert corpora are discussed in this part, it should be pointed out that (mostly small) learner corpora are frequently used in the ESP field, since teachers are concerned with the production of their students in contexts of specialization. While general corpora have proven to be most effective for the study of the structure and use of language, specialized corpora which focus on specific genres are required when exploring language in specific academic and professional settings (Connor and Upton 2004a). According to Flowerdew (2004), specialized smaller corpora offer more advantages than general corpora from a methodological perspective because they provide more contextual information (i.e. the communicative situation) than larger corpora. When complete texts are included, the implementation of top-down analyses of the textual and generic features present in the texts is made feasible.

Similarly, genre analysis clearly benefits from the use of specialized corpora, which help to grasp more accurately the function and use of language in genre. In this sense, corpus linguistics reveals itself as an essential and indispensable framework which, combined with genre analysis (Swales 1990; Bhatia 1993), may provide new insights and ultimately help ‘to improve the training of novice writers and to encourage the development of better and more effective [texts]’. (Connor and Upton 2004b: 254).

The first two chapters in the part Corpora and English for Specific Purposes study the use of written and spoken academic English corpora. Annelie Ädel (Chapter 3) provides a thorough review of the challenges that lie ahead in the use of corpus for the teaching of academic writing. She discusses the scarce attention paid to the potential of corpora in the context of writing instruction. As she rightly states (Ädel, this volume: 41): at this point in time, it takes a corpus linguist to offer a corpus-based writing class.’ To alleviate this situation, she presents seven different challenges involved in using corpus-based approaches in teaching writing in English for Academic Purposes (EAP) settings. Among these challenges we find the lack of corpus availability; the difficulty of finding what users are looking for, where and how, without getting lost in large amounts of data; how to evaluate and present corpus patterns to language learners; how to manage decontextualized data; and how to connect surface forms to meaning.

English subject curricula should take into account language aspects that go beyond linguistic features to introduce real language into the classroom. Thus, in Chapter 4, Begoña Bellés and Mari Carmen Campoy explore the uses of the phrase ‘I feel’ and its variants in contiguous and
non-contiguous collocational patterns by analysing MICASE data, indicating lexico-grammatical and textual features that should be taken into account in the teaching of communicative functions in spoken academic discourse at a wide range of linguistic levels. They suggest that these findings should be included in class so as to contribute to raise the student’s awareness of the connection between grammatical, sociolinguistic and pragmatic uses of phraseological items of the English language. The analysis of ‘I feel’ is a good example of how the MICASE search engine may aid in teaching the use of a stance verb by highlighting it as marked in terms of uneven distribution among genres, speech event interactivity rate and in its use among different genders. This is a complex teaching approach to the analysis and understanding of modality devices that teachers may only carry out thanks to the annotation of speaker and speech-event categories that the corpus search interface makes possible.

The following four chapters in this section (by Winnie Cheng, María José Luzón, Belinda Crawford, and Maria Georgieva and Lilyana Grozdanova) deal with corpora and English for Professional Purposes (EPP). An interesting feature of learner corpora in this context is that text or speech production on the part of learners does not usually coincide with the text types and genres collected in native speaker corpora. In the area of EPP, however, this situation is changing. EPP teachers are now gradually becoming more engaged in trying to get their students to produce texts based on language use situations in which they might find themselves in their future as professionals in a specific area of work.

Cheng (Chapter 5) shows how ConcGram© (Greaves 2009) may be used to elicit data from a corpus representing the English language of the engineering sector in Hong Kong, and discusses how the results may be used to deal for instance with the aboutness of the text and to help EPP students to learn the language used in their profession. Regarding the use of NS and NNS corpora for ESP teaching (see Gavioli 2005 for ESP corpora designed with teaching purposes rather than for language description), characteristic discourse moves may be studied by learners so that they become aware of those common expressions that are typical of the genre under analysis within the wider perspective of move sequencing.

Luzón (Chapter 6) studies the misuse or atypical use of organizational items in a small learner corpus in contrast to the information gathered from the BNC corpus. The problematic areas found in Luzón’s study include errors regarding the word class, meaning, or function of an item and its position in a sentence, as well as atypical or incorrect use of (sometimes inexistent) lexical bundles and genre phraseology. Errors discussed
in Luzón include those involving signalling nouns and their use to create cohesive relations across-clause level. Likewise, problems regarding the use of informal or oral discourse in a formal context are brought to light. In this chapter it is made clear that in order to design effective teaching materials it is essential that both native speaker and learner corpora should be brought together to better understand learner’s needs and problematic areas in order to identify language patterns used by learners which clearly differ from those used by experts.

In Chapter 7, Crawford introduces a spoken business corpus and derived classroom activities that may improve ESP materials through corpus-based pedagogical applications. Drawing on a small specialized corpus, the author explores key business English lexis and demonstrates that corpus-based activities can help students better understand content lectures in English. This is vital for the learners’ success not only in their academic studies but also in their future careers. In this way, Crawford (this volume, 104) contributes to ‘bridging the gap between ESP research and ESP pedagogy’.

The last chapter in this part, by Georgieva and Grozdanova, pays special attention to English as an International Language (EIL) or English as a Lingua Franca (ELF) corpora (Seidlhofer 2005; Mauranen 2007). EIL/ELF corpora are particularly focused on the production of native-like speakers in academic and professional contexts. For the majority of ESP learners, competent professional communication is one of the highest motivations to learn a language. Georgieva and Grozdanova intend to answer a different set of questions, such as which strategies participants in intercultural communicative encounters use to overcome differences in the process of communicating with other speakers; or which are the most widely used patterns that come up in order to communicate successfully.

1.2 Learner Corpora and Corpus-Informed Teaching Materials

The corpora explored in this part may be termed pedagogic corpora (Hunston 2002) or (E)LT discourse corpora,¹ in a similar fashion to EIL corpora. They include the language used in classroom or in formal teaching and learning contexts and situations (exams, office tutorials, etc.) and may take into account teacher-learner relationship patterns. A comprehensive example of this kind of corpora is the T2K-SWAL corpus designed to test to which extent the language of ESL/EFL materials and assessment instruments represents ‘real’ English language (Drescher 2007; García 2007). This group
would also include English Language Teaching (ELT) materials corpora, in the sense that textbooks, for instance, are meant to represent NS production as a model for the language learner (Römer 2005; Amador-Moreno et al. 2006; Cheng 2007). Authenticity of the written/spoken texts is questioned here in terms of the language used, the text types provided and the authenticity of tasks. Corpus Linguistics has a lot to say in the assessment/improvement of the aforementioned levels of authenticity. Corpora based on the interaction between teachers and learners which should be considered EPP corpora would fall into this category. Examples of analysis of this interaction may be seen in the MICASE corpus (Csomay 2007), or the POTTI corpus (Farr 2007) and also in O’Keefe et al. (2007: 220–243).

In this volume, the part devoted to ELT corpora focuses on three main dimensions: the first one deals with the compilation and exploitation of learner corpora; the second explores error analysis using learner corpora and comparable native speaker corpora; and the third has to do with the use of corpora to create teaching materials.

The compilation and use of corpora as a means to enhance language learning practices takes us to the issue of criteria in corpus compilation which determine the end product and how and by whom it may be used afterwards (Luzón et al. 2007: 4–6). Among these, there are at least three essential criteria that affect corpus-based language learning and teaching: (1) the purpose and principles behind the compilation of the corpus, (2) its availability, not only for the researcher but also for materials writers, teachers and learners and (3) the use of various resources in multimodal corpora. In this sense, as may be seen in the articles collected in Ghadessy et al. (2001), it is a well-known fact that a good number of teachers prefer the use of small \textit{ad hoc} corpora that have been designed with a very specific aim in mind and addressed to a particular group of learners. There are two obvious reasons for this: one is that, given the opportunity, teachers would not avoid the possibility of tailor-made resources; the other is that, in most cases, small \textit{ad hoc} corpora are easier to handle in the classroom.

If we consider the issue of corpus compilation purposes, another interesting feature stands out: how the texts are obtained, i.e. the compilation methodology. Thus, we think that an important point when dealing with corpus-based methodologies is that learner corpora follow a task-based instead of a text-type based approach in their compilation and database organization. This takes us to the subject of how learner corpora differ from corpora with other speaker profiles. In learner corpus compilation, an important debate revolves around the kind of task selected to elicit learner
language production, and the extent to which the elicited language may be seen as authentic. In this sense, it is important to bear in mind that any chosen task for the learners is not going to be considered as natural as those performed by native speakers since the former are produced in a more or less imposing learning situation where fully spontaneous speech may not be attained, though it may be argued, as in Sylvie De Cock’s chapter (Chapter 9), that the learning situation is in fact a real situation for learners.

In her chapter, De Cock extensively reviews the use of spoken learner corpora in ELT. She discusses the two fundamental aspects in learner corpora: learner variables and task variables. Learner variables pose a number of questions regarding the complexity of the description of speaker profiles and of the compilation of speaker production corpora where speakers follow the same procedures and belong to a similar learning profile. Task variables largely influence not only what may be done with the corpus in question but also the possibility of research replications in subsequent investigation. For the creation and analysis of oral tasks, communication problems arising from inability to convey a message are one of the main concerns when querying corpora and they constitute a central issue when designing pedagogically relevant materials.

Moreover, De Cock complains about the scarce availability of materials derived from spoken corpora, which are also still in its infancy regarding classroom exploitation. Direct and indirect use of spoken learner corpora requires participation on the teachers’ part that could at this stage perhaps only be carried out if the teacher is a corpus linguist or is trained specifically to deal with such corpora, since spoken corpora are difficult to handle at least in depth or to obtain as many benefits as possible on the part of both learners and teachers.

In the second chapter of this section, Julia Lavid, Jorge Arús and Juan Rafael Zamorano explain details about the compilation and exploitation of a small bidirectional corpus of written texts. The texts in their online corpus include originals and their translations in English and Spanish, and allow for the analysis of individual texts as well as for ‘whole-corpus reading’. In an effort to guide teachers and learners, the authors also include other tasks which would fit into what is called direct use of corpora, designing possible hands-on tasks as part of their corpus-based materials.

Regarding the use of corpora to analyse learner output, Chapters 11, 12 and 13 (by Rafael Alejo, Mª Ángeles Andreu et al. and Amaya Mendikoetxea et al., respectively) explicitly deal with corpus-based error analysis and learners’ non-prototypical use of English. Many studies analysing learner
Corpus-Based Approaches to English Language Teaching
corpora focus to a large extent on language proficiency and on possible
ersrors in a set-up task. In Chapter 11, Alejo explores the Spanish and
Swedish components of the ICLE corpus and the Written School and
University Essays from the BNC to compare the use of the particle ‘out’ in
both corpora in terms of over- and underuse, prototypicality, avoidance and
erroneous use of this particle. Similarly, Andreu et al. (Chapter 12) analyse
written production of EFL students in an error-annotated multilingual
corpus of students learning English, Spanish, French and German as a
foreign language, and also Catalan, as a first, second or foreign language.
Comparable and parallel multilingual corpora incorporate the production
of speakers (NS or NNS) whose mother tongue may represent two or more
languages. They are most common in corpus-based translation studies. The
possibilities are varied: researchers, teachers and students may be using
comparable and/or parallel corpora in two or more languages to analyse
possible translations and/or to check on a specific language issue. Other
multilingual corpora discussed in this volume may be found in other parts
(see Lavid et al.; Alcaraz et al.; Guzmán and Alcón).

Mendikoetxea et al. (Chapter 13) aim at the development of teaching
materials drawing on a database of learner errors extracted from a corpus
of essays written by Spanish learners of English at university level in order
to identify problematic areas and to develop relevant pedagogical materials,
thus improving curriculum design. Their project (INTELLeNG) combines
contrastive analysis (CA) and error analysis (EA). Despite advocating
for the use of learner corpora, the authors highlight the benefits of the
combination of learner and native corpora for the elaboration of teaching
materials and curriculum design as part of classroom methodology aimed
at fostering students’ language awareness and, ultimately, their language
proficiency.

The last three chapters of this part deal with the creation of corpus-
informed language teaching materials taking into account lexicography,
grammar and representativeness in language learning. Learner corpora may
be used to obtain feedback for the improvement of existing pedagogical
materials. In this area, corpus-based updating and improvement of peda-
gogical dictionaries is one of the most widely exploited fields of research.
Grammar and textbook design are now also receiving more attention in the
field of indirect corpus applications. Cheng (2007) and Römer (2005) are
examples of how differences between actual language use and textbook
language may be tackled by means of corpus analysis.

In Chapter 14, Sylvie De Cock and Magali Paquot discuss the design of
corpus-based information in dictionaries that are meant to aid learner
language production. They focus on the work carried out in the *Macmillan English Dictionary for Advanced Learners* to describe how the International Corpus of Learner English (with writings of learners from 16 different countries) is used, together with information drawn from a 15 million-word corpus of academic English in order to provide improved information on those areas where difficulty was detected in the learner corpus. Thus, corpus information is an added value in the form of ‘Get it right’ boxes, grammar sections and academic writing sections, increasing the dictionary’s productive use potential.

Chapter 15 also deals specifically with corpus-informed teaching and learning materials. Here, Tom Rankin analyses adverb placement in an advanced learner corpus suggesting ways to improve grammar teaching materials. Adverb syntax is a particularly problematic area for EFL learners but, paradoxically, it has been neglected in most grammar textbooks. Rankin contends that specific discourse and pragmatic contexts must be taken into account when teaching adverb placement and suggests that corpus data can inform the selection and sequencing of materials and ‘provide practical help in choosing which type of semantic and syntactic features prove most problematic for the learners and should therefore be included in teaching examples and exercises’ (Rankin, this volume: 305).

Finally, the issue of representativeness in corpora use and compilation is discussed by Izaskun Elorza and Blanca García-Riaza in Chapter 16. These authors tackle the question of how the compilation of a successful pedagogical corpus of written academic texts should be done in terms of size, topic, authenticity and representativeness. The focus remains on the texts chosen for the corpus, since learners will take them as a model of the language used for ‘real’ communication. The authors suggest that the use of a specific (pedagogic) corpus can influence the definition of the model of language to be used in the classroom. As the authors indicate (Elorza and Riaza, this volume: 221),

> when dealing with the corpus compilation of the written input we cannot ignore the great variety of the texts used in higher education courses. The need for using texts from different types seems to impede the very possibility of compiling a representative corpus in terms of typological representativeness.

Thus, they study wordlist statistics, rank and frequency of word types in relation to text length, completeness and representativeness and compare and contrast data to the first hundred most frequent words in the BNC corpus.
1.3 Multimodality: Corpus Tools and Language Processing Technology

The development of corpus tools and the integration of different modes of communication in corpora are key issues in the use of corpora for learning purposes. Also, CD and online availability allow both learner and teacher to use corpus resources at ease. Together with this availability is the issue of user-friendliness in the design of both corpus and corpus tools. The fact that most educational institutions have access to the internet has promoted the use of the web as corpus (Kilgarriff and Grefenstette 2003; Sharoff 2006) in the making of self-compiled *ad hoc* corpora, since educators worldwide find it easy to download the exact text types they need to use in the classroom and make them part of a corpus in a do-it-yourself fashion (e.g. *CorpusBuilder* in *SketchEngine*). In this sense, web as corpus research facilitates the study of multimodal features through the use of corpora. Moreover, the development of customized corpora such as ACORN (the Aston Corpus Network) and its focus on, and open access to, corpus and corpus output materials show how corpora are increasingly present in today’s educational institutions.

Some CD and online language learning packages also include corpora as part of their components (see for instance the Virtual Language Centre, Hong Kong Polytechnic at http://www.edict.com.hk/vlc/ and its *WebConcordancer*), so learners may play around with several search routes which allow for teacher work on various language proficiency levels. Furthermore, with the combination of different discourse modes in multimodal corpora, learners may develop all four competences.

We would also like to point out the advances that have been made since Tim Jones’ pioneering work in DDL (Data Driven Learning), when most research was based on concordance and collocation data. The future that lies ahead regarding corpus tools that may be used by learners and teachers alike is more complex and exciting than ever. First, the availability of a wide range of corpora, which may be operated through diverse corpus tools, enables teachers to design a wide range of materials and tasks for the classroom. The creation of corpora such as MICASE including speech events, speaker status and academic position, speaker level or interactivity rating of the event, makes it possible to go beyond the word and its lexico-grammatical patterns into other discourse levels. Secondly, a surge for pedagogic annotation and annotation tools (Braun 2006; Alcaraz et al., this volume) reflects the interest of teachers and researchers alike to use annotated corpora in the classroom and in the creation of language teaching...
materials. New corpus tools such as SketchEngine and its Word Sketch automatically provide the user with a complete collocational and grammatical pattern of searched words and phrases; others, like the Word Sketch differences, show lexical contrasts between two selected words in terms of their collocates.

The possibility to study word association and the combination of genre and keyword analysis (Scott and Tribble 2006) by means of tools such as WordSmith Tools gave corpus studies a wider dimension. The development of new tools in this direction may be seen in ConcGram® (Greaves 2009), a programme which determines the phraseological profile of the language contained in a specific corpus. As described in Cheng et al. (2006), many word associations do not occur in one fixed grammatical pattern so, taking this into account, ConcGram® develops information based on non-contiguous sequences of associated words (Cheng et al. 2006: 414):

The development of the notion of a concgram challenges the current view about word co-occurrences that underpins the KWIC display (. . .) word associations become the focus of attention, and a ‘node’ is not the ‘sun’ around which collocates orbit in a subordinate relationship.

As can be observed, tool and multimodality play an active role in the development of corpus-based approaches to ELT. The final part of this volume examines availability and multimodality in corpora within the language teaching context, and presents several new devices for corpus processing, introducing tools such as a query program for parallel corpora or a tool for implementing pedagogical annotation. The chapters discuss the opportunities and challenges that multilayered and multimodal corpora may pose to corpus linguistic investigation in ELT.

More specifically, José María Alcaraz et al. (Chapter 17) show a tool that allows annotation for any language and explain how the seven language corpora in the SACODEYL project can be annotated with the same tool, thus providing useful resources in the pedagogical design and analysis of classroom material. In Chapter 18, Josep Roderic Guzmán and Eva Alcón use two corpora made up of TV series in English which are translated into Catalan and Spanish, and narrative works where English is the language of the original texts, also translated into Catalan. They explain how these corpora may be approached by means of the AlfraCOVALT tool in order to design tasks to cater for the use of requests in English, and how to apply the data provided by the corpora to the creation of activities for translation students.
In Chapter 19, Inmaculada Fortanet and Mercedes Querol present their experience in the compilation of a multimodal video corpus recorded and edited for its application to a teacher training course for lecturing in English at Universitat Jaume I. These authors offer an example of the type of tagging or classification of speech events that can be done in video corpora, which can later assist the teaching of pragmatics, grammar and/or vocabulary. They advocate for multimodal corpus analysis, stating that when teaching spoken academic discourse by means of corpus-based learning, corpora transcripts do not always provide enough information about the real situation, lacking of general context and background. They conclude that language is accompanied by prosodic features such as intonation, accent, or stress and kinesics which cannot be exclusively analysed from a transcript.

If there is a promising future in corpus studies in the ELT field, it is that of multimodal corpora and the tools developed to support them. The study and analysis of multimodal corpora could be understood as a critical rethinking and reformulation of the relationship between text and society (Baldry and Thibault 2006: 2). This provides researchers with other language, social and cultural aspects not gathered or embedded within a linear approach or analysis. A not-single theoretical framework such as the analysis of multimodal corpora, can in fact adequately describe the very different semiotic systems (language, music, picture, movement, etc.). By analysing multimodal corpora, researchers do not only aim at the study of plain texts or transcripts, but other modes of discourse are taken into consideration and seen as a unique whole. How all these modes of discourse are interrelated or not, structured, organized and presented, can be studied by means of multimodal corpora. As Baldry and Thibault (2006: 3) point out, ‘text users’ knowledge of culture and society interact with the internal features of text’s organization during the making and interpreting of texts.’ It should be added at this point that we, as linguists, understand text not only as a written mode of discourse. With multimodal corpus analysis we are not limited to text analysis; there are many other resources that can be used to create or support texts, a phenomenon which has been referred to as resource integration principle (Baldry and Thibault 2006: 4).

Many are the ideas to be drawn from this volume. However, we would like to underscore two central issues. One is the fact that research-oriented corpus tools have still a lot to say in indirect corpus applications, that is, on what and when to teach. This is more so for learner corpora and for spoken corpora, due to the fact that these are the most difficult corpora to compile but are also, or should be, more productive in terms of providing data that
may be applied to the classroom in a satisfactory way. The same is true about teaching-oriented corpus tools and their role in direct corpus applications, since the development of these tools together with the analysis of teacher and learner needs will undoubtedly lead to a more active participation of both teachers and learners in the corpus-based learning process, that is, in how we may teach and learn a language. In this sense, we can remain assured that the future of corpus linguistics and language teaching will go hand in hand to provide valuable and much needed pedagogical applications, to improve teaching materials and course syllabi, and ultimately to respond to the needs of both teachers and learners.

A second, final issue concerns the reflection made around concepts introducing the prefix ‘multi’ in combinations such as ‘multilayered’, ‘multimodal’, ‘multipurpose’, ‘multilingual’, ‘multiple tools’, ‘multiple annotation’, etc. We would like to take the ‘multi-combinations’ terms used throughout this volume as an emblem towards the new and exciting challenges that the new corpora and updates of the old ones bring on to the stage for corpus linguistics and ELT.

Notes

1 The term pedagogical corpora might imply study and evaluation of that discourse as pedagogical, without questioning the efficiency of that discourse in learning contexts. Use of teacher and teaching materials corpora may sometimes reveal a bigger or lesser degree of pedagogical inadequacy.

References

Corpus-Based Approaches to English Language Teaching


Chapter 2

Using General and Specialized Corpora in English Language Teaching: Past, Present and Future

Ute Römer
University of Michigan

2.1 Introduction: Corpus Linguistics and Language Teaching

Over the past 25 years, corpora, corpus tools and corpus evidence have not only been used as a basis for linguistic research but also in the teaching and learning of languages. Tim Johns’s data-driven learning (DDL), Dieter Mindt’s empirical grammar research, and John Sinclair’s work with COBUILD can be considered particularly groundbreaking developments in the field of English corpus linguistics and language pedagogy in the 1980s (see Mindt 1981, 1987; Johns 1986, 1991; Sinclair 1987, 1991).

Nowadays, more and more researchers and practitioners treasure what corpus linguistics has to offer to language pedagogy, and the impressive number of recently published monographs and edited collections on the topic clearly indicate the growing popularity of pedagogical corpora use and the need for research in this area (see, for example, Aston 2001; Granger et al. 2002; Sinclair 2004a; Römer 2005; Ädel 2006; Braun et al. 2006; Gavioli 2006; Kettemann and Marko 2006; Scott and Tribble 2006; Campoy and Luzón 2007; and the proceedings of the first six events in the TaLC (Teaching and Language Corpora) series: Aston et al. 2004; Botley et al. 1996; Burnard and McEnery 2000; Hidalgo et al. 2007; Kettemann and Marko 2002; Wichmann et al. eds. 1997).1

I would, however, still be hesitant to say that corpora and corpus tools have after all fully ‘arrived’ on the pedagogical landscape. The practice of ELT (English Language Teaching) to date, at least, seems to be largely unaffected by the advances of corpus research, and comparatively few teachers and learners know about the availability of useful resources and get their hands on corpus computers or concordancers themselves (see