UTE RÖMER

Pedagogical Applications of Corpora:
Some Reflections on the Current Scope and a Wish List for Future Developments

Abstract: Despite the progress that has been made in the field of corpus linguistics and language teaching, the practice of ELT has so far been largely unaffected by the advances of corpus research, and corpora and concordances are hardly used in the German EFL classroom. This article aims to take stock of developments in this area and formulates tasks for the future of pedagogical corpus applications. After a brief overview of relevant corpus resources and a discussion of their availability, I will give an introduction to some direct and indirect approaches to using corpora in English language learning and teaching. The paper will then explore some current limitations of pedagogical corpus work and close with comments on a few desiderata in research and practice.

1. Introduction

In the context of language learning and teaching, corpora and corpus evidence have been used for more than twenty years now. John Sinclair's work with COBUILD,[1] Tim Johns's data-driven learning (DDL), and Dieter Mindt's empirical grammar research can be considered particularly groundbreaking developments in this field in the 1980s (cf. Sinclair ed. 1987 and 1991, Johns 1986 and 1991, Mindt 1981 and 1987). Nowadays, an increasing number of researchers and practitioners treasure what corpus linguistics has to offer to language pedagogy and regard corpora as useful tools in their everyday work.

However, despite the progress that has unquestionably been made in the past two or three decades, I would still be hesitant to say that corpora have after all fully 'arrived' on the pedagogical landscape. Most recent research shows that the English language teaching practice, at least in Germany, has been largely unaffected by the developments in corpus linguistics (cf. Mukherjee 2004), that corpora and corpus methods play a very minor role in EFL initial teacher training at universities (cf. Breyer 2005), and that the currently used pedagogical language descriptions differ considerably from actual language use as captured in corpora

---

[1]   COBUILD is an acronym which stands for COllins Birmingham University International Language Database. For a detailed account of the COBUILD project see Sinclair ed. (1987).

(cf. Römer 2005). Despite the obvious and recognised strengths of corpus use in a pedagogical context, e.g. that corpora highlight what lexical items and collocations are typical in the language, and that they provide us with large amounts of natural language examples (cf. Hunston 2002, 20; Römer 2005, 3 and 276), it seems that there is still a strong resistance towards corpora from the side of students, teachers, and materials writers.

The aim of the present article is to briefly review what has actually been achieved so far in the field of corpus linguistics and language teaching (CL & LT), and to discuss what should be done that has not yet been accomplished. In the following, I will first describe some useful corpus resources and provide a short overview of the major types of pedagogical corpus applications. The main focus of the paper will then be on the current limitations of pedagogical corpus work and on some desiderata in research and practice. I will devise a CL & LT wish list and conclude with a few recommendations for a future agenda of research and practice in the field.

2. The availability of useful resources

A quick web search for "corpora" or "corpus linguistics" finds websites with long lists of language corpora, a wealth of information about research activities and publications, and links to a number of software packages for corpus analysis. From the wide range of available resources, I have picked just a few that I consider useful for the direct and indirect application (cf. Sections 3.1 and 3.2) in language learning and teaching, and grouped them into four categories: (i) corpora on CD-ROM, (ii) corpora on the web, (iii) concordance packages, and (iv) corpus-based reference works and teaching materials.

A number of corpora of different types of spoken and written English can be purchased on CD-ROM and are then installed on a personal computer, for use by a teacher or researcher, or on a local computer network, e.g. for use with a group of learners. These corpora either come with their own specific software for corpus analysis, or they can be used with other freely or commercially available concordance packages (see below). Worth mentioning are the corpus collections distributed by the International Computer Archive of Modern and Medieval English (ICAME), the members of the British National Corpus family (BNC version 1.0, BNC Sampler, BNC World, and BNC Baby), and the British component of the International Corpus of English (ICE-GB).[2] Particularly useful for pedagogically oriented research are, in addition to these corpora of native-speaker English, the eleven parts of the International Corpus of Learner English (ICLE), which consist of the written production of EFL learners with eleven different first language backgrounds (e.g. French, German, and Swedish).[3] Other corpora that are available

---

[2]   Relevant websites with information about all resources mentioned in this section are listed in Appendix 1.
[3]   For overviews of ICLE and other learner corpora and their research potential, see Barlow (2005), Granger (2004), Nesselhauf (2004), and Pravec (2002).

on CD-ROM include the first release of the American National Corpus (ANC) of spoken and written American English, and the Santa Barbara Corpus of Spoken American English (SBCSAE). Most of the corpora mentioned here can be labelled "small" by today's standards (not the 100 million word BNC, though), but I would still consider them very useful repositories of natural language and a valuable basis for the creation of concordance exercises and other teaching resources.

While the use of CD-ROM corpora normally requires the payment of a license fee, access to corpora on the WWW is in general free of charge. Web interfaces provide access to corpus data, and allow for simple word and pattern searches to retrieve KWIC (KeyWord In Context) concordances (sometimes with a limited maximum output of 40 or 50 lines) or lists of collocations. Recommendable for searches on general English are the Collins Concordance and Collocations Sampler (based on a 56 million word subset of the Bank of English), the BNC World online service, and the VIEW interface of the BNC (VIEW = Variation In English Words and phrases). A more specialized type of English can be accessed through the search page of the Michigan Corpus of Academic Spoken English (MICASE). Recently, efforts are being made to domesticate the web so that it can serve as an immensely large corpus. Pioneers in this field are Antoinette Renouf and her fellow researchers in the WebCorp team (URLs in Appendix 1).[4]

Concerning the types of analytic software for use with any of the CD-ROM corpora, researchers, teachers, and students can again choose between some basic freeware tools and licensed packages which offer a wider range of features for corpus access and analysis. Examples of the former kind are Chris Greaves's ConcApp and Laurence Anthony's AntConc. A lot more sophisticated and more widely used among linguists, however, are two packages of the latter type: WordSmith Tools (Mike Scott, OUP) and MonoConc Pro (Michael Barlow, Athelstan). In addition to the central concordance function, both packages provide tools to extract word-clusters, to calculate collocations, and to determine the distribution of words and phrases across texts.[5]

Some reference works and teaching materials now incorporate corpus evidence and contain authentic instead of invented examples. Based on real English and compiled with the needs of the language learner in mind, the materials in the Collins COBUILD series (dictionaries, grammar and usage books, concordance samplers) are highly recommendable and certainly offer teachers and learners more reliable information about the language than any of the more traditional reference grammars or non-corpus-based dictionaries. Following the example of COBUILD, other corpus-based dictionaries (often on CD-ROM), grammars, usage handbooks, and collocations dictionaries have been made available in the

---

[4]   The 'web-as-corpus' problem was also in the focus of attention at the recent "Corpus Linguistics 2005" conference (University of Birmingham, 15-17 July 2005) and at a symposium on "Corpus Linguistics – Perspectives for the Future" (University of Heidelberg, 21-23 October 2004; cf. Hundt, Nesselhauf and Biewer eds. forthcoming).

[5]   For an account of the basic steps in corpus analysis (using WordSmith Tools), the reader is referred to Römer (forthcoming).

past few years (see Appendix 2). Also very useful, especially for work with advanced learners, are publications that offer concordance exercises designed by corpus linguists, e.g. *Concordances in the Classroom* (Tribble and Jones 1997) or *Reading Concordances* (Sinclair 2003). EFL textbook series have long remained unaffected by the so-called 'corpus revolution', but fortunately, the first attempts are now being made to produce course books which draw on corpus research and are fully based on real-life data (see references to Barlow & Burdine forthcoming and McCarthy et al. 2005 in Appendix 2).

3. A brief overview of pedagogical corpus applications

Let us now turn from corpus resources to some of their most important pedagogical applications. A useful distinction that can be made in this context is the one between direct and indirect approaches to using corpora in language teaching.

3.1 The direct approach

To follow a direct approach to pedagogical corpora use means that teachers and learners get their hands on corpus data themselves, instead of having to rely on the researcher as mediator or provider of corpus-based materials. Tim Johns, the 'father' of this direct, so-called 'data-driven learning' (DDL), approach, suggests to "confront the learner as directly as possible with the data, and to make the learner a linguistic researcher" (Johns 2002, 108). Johns's motto for this inductive learning approach in which learners work with concordances and consult corpora in an exploratory way is "'Every student a Sherlock Holmes!'" (ibid.). In a similar vein, though with a shift from 'learning as research' to 'learning as discovery', Bernardini (2002, 165) talks about the positive effects of "corpus-aided discovery learning" and describes corpora as "rich sources of autonomous learning activities of a serendipitous kind" (ibid.; cf. also Bernardini 2000).

   Clearly a major advantage of DDL and related approaches, both for the teacher and for the learner, is the immediate accessibility of authoritative information about what is acceptable in the language and about how certain words or phrases are actually used by competent speakers of English. As fittingly stated by Barnbrook (1996, 140), "students can derive the information they need directly from the language, as though the computer were a tireless native-speaker informant, with rather greater potential knowledge of the language than the average native speaker." As an alternative to the 'direct corpus access' strategy, teachers can opt for a somewhat more controlled way of DDL and provide learners with (raw or filtered) concordances, or with concordance exercises which highlight a particular language problem. For instance, a teacher who wants to emphasize the semantic and phraseological differences between the verbs "speak" and "talk" and who has access to a corpus could easily create a corpus-based 'fill-

What is the missing word in each of the following sentences – 'speak' or 'talk'?

```
I can only _____ for myself; I can't _____ for you of course.
Are you able to _____ English fluently?
And then at the end of it we'll _____ it through.
I managed to put her off that idea, managed to _____ her out of that.
So you're free to _____ your mind.
Can you _____ up a bit? _____ up a bit Belinda!
I will _____ to David about it as well.
I'd like to _____ very strongly in favour of it.
And I'm sure I _____ on behalf of you all when I say thank you very
much indeed to Robin.
Men tend to _____ like that, don't they?
```

Figure 1: An example of a possible DDL exercise – 'speak' vs. 'talk'

the-gap' exercise of the type presented in Figure 1.[6] Depending on the learner's needs, such exercises can be based on all sorts of corpora, e.g. of a special text type or a particular regional variety of English.

3.2 The indirect approach

While the direct approach features corpus use by teachers and learners and introduces corpora and concordances to the classroom, the indirect approach is researcher-focused and centres on corpus evidence and the impact it can have on syllabus design or teaching materials. Linguists who follow this approach carry out research at the interface of corpus linguistics and language teaching, with the aim to provide better descriptions of 'used' language and with the question in mind how corpus work can contribute to an improvement of language teaching and help to make life easier for the learner.

We can distinguish different types of indirect pedagogical corpus applications. One is to investigate learner corpora and compare the language patterns found in the spoken or written production of learners with the patterns found in a comparable native-speaker corpus. Such studies usually start out from language features that are known to be common problem areas for learners, for instance modal verbs (cf. Aijmer 2002), tenses (cf. Granger 1999), connectors (cf. Granger and Tyson 1996), or verb-noun collocations (cf. Nesselhauf 2005). The same is also true for analyses which compare the actual use of problematic lexical-grammatical items with the presentation of the same items in EFL/ESL teaching materials (mainly coursebooks and learner's grammars). Among the topics that have been examined in that kind of corpus-coursebook comparisons are irregular verbs (Grabowski and Mindt 1995), future time expressions (Mindt 1987 and 1997), linking adverbials (Conrad 2004),

---

[6]   For further examples of DDL exercises, see Tim Johns's and Passapong Sripicharn's websites (URLs listed in Appendix 1).

modal verbs (Römer 2004a), *if*-clauses (Römer 2004b), and the present perfect (Schlüter 2002). For all items investigated, researchers found considerable mismatches between naturally-occurring English and the English that is put forward as a model in pedagogical descriptions. What such findings clearly indicate is the need for an improvement of teaching materials, so as to bring what is taught in accordance with what is commonly used in the language.[7]

Yet another type of studies examines parallel or translation corpora, i.e. collections of source texts in one language that are aligned with their translations into another (target) language, and retrieves pedagogically relevant findings from contrastive language data. Since corpus-based contrastive analyses, among other things, provide insights about translation equivalents or the lack thereof, their results can be used to inform the compilation of bilingual dictionaries and learner's dictionaries. Knowledge about language items which cause problems to translators may also be of use for teachers and materials writers.[8]

4. Scope for development: A CL & LT wish list

In the previous sections of this article, I have tried to give a brief account of the state-of-play in CL & LT, mainly concerning the availability of resources and the kind of research that has been carried out so far. I will now critically approach the topic and focus on the still existing limits of pedagogical corpus applications. In other words, now that we know what is currently available and what is done with the existing resources, the question is 'What else should be available and done in the area of CL & LT?' A wish list for future developments will serve to answer this question. It addresses the following central issues: (i) resources, (ii) research activities, and (iii) 'missionary work'.

4.1 Resources

From what has been said above about the availability of a range of corpus resources (cf. Section 2), it may seem that everything we need is right there, just waiting to be used. However, we notice a lack of some types of corpora that are considered to be particularly useful in a pedagogical context.

In general, spoken English corpora are a lot more difficult to get hold of than corpora of written English (which are usually also much larger than their spoken counterparts, for apparent economical reasons). This is rather unfortunate, especially because language teaching curricula normally emphasise speech over writing and call for materials that help to foster the learner's communicative competence. It would hence be good to have access to more (and richer) sources of spoken language which could inform or serve as a model of teaching materials and pedagogical reference

---

[7] For a large-scale study on the use of progressive verb forms in real English and 'school' English which exemplifies the pedagogical potential of corpora, see Römer (2005).

[8] For reports on parallel corpus studies that include discussions of pedagogical implications, the reader is referred to Barlow (2000), Claridge (2002), or Schmied (1998).

works. I am here not only thinking of general native-speaker English speech corpora but also of spoken learner corpora,[9] 'spoken-type' parallel corpora,[10] and classroom discourse corpora (going back to Sinclair and Coulthard's 1975 study on teacher-pupil interaction). Besides, the compilation of more and larger collections of EFL coursebook language, such as GEFL TC (the German English as a Foreign Language Textbook Corpus, cf. Römer 2004b and 2005) or the TeMa corpus (cf. Gouverneur and Meunier 2004; Meunier 2004), would enable us to carry out further comparative analyses of real English and 'school' English.

With respect to concordance analysis packages, we have to consider whether the tools that are currently available are easy enough to use for learners and teachers who, obviously, have not had the same training as the corpus-linguistic researcher. Another item on my resources wish list would therefore be an easy-to-use, appealing concordance program (or a suite of programs) that teachers and learners would be willing and able to work with. Even if this may sound a difficult problem to solve, it would be very useful if, with just one or two mouse clicks, some kind of advanced structuring of concordances (more than just an alphabetical sorting of the context) could be provided, for instance to ensure quick access to relevant examples that show typical collocations, or to extract common patterns and highlight the different senses of a polysemous item. What I would also like to see more of in the future, are coursebook series (like McCarthy et al.'s *Touchstone* or Barlow and Burdine's *CorpusLAB* books) that are entirely corpus-based and that are tailored to the needs of particular groups of learners, e.g. learners with different first languages and of different proficiency levels. Reference works, like learner's grammars or usage handbooks, could also be improved if they were informed by results from contrastive and learner corpora research or from real-English-vs.-'school'-English comparisons. On a similar note, Granger recently said that she is impatiently waiting for a learner corpus-informed EFL grammar.[11]

4.2 Research activities

The various activities of a growing number of researchers in the TaLC (Teaching and Language Corpora) movement indicate that corpus work can indeed have an impact on language pedagogy.[12] Insights provided by contrastive and learner

---

[9]   A learner corpus of spoken English produced by German learners is now being compiled by a research team around Joybrato Mukherjee at the University of Giessen as part of the LINDSEI project (LINDSEI = Louvain International Database of Spoken English Interlanguage, cf. http://cecl.fltr.ucl.ac.be/Cecl-Projects/Lindsei/lindsei.htm).

[10]   As noted in Römer (2005, 296) the compilation of a parallel speech corpus obviously causes some technical problems. A first step towards compiling 'spoken-type' parallel corpora could, however, be the extraction of original and translated subtitles from movie DVDs.

[11]   This wish was expressed by Sylviane Granger during a colloquium on "Corpora and language teaching: the state-of-play and future perspectives in research and practice" at the "Corpus Linguistics 2005" conference in Birmingham, UK, July 15th 2005.

[12]   See the proceedings volumes of the first five conferences in the TaLC series: Wichmann et al. eds. (1997), Botley et al. eds. (1996), Burnard and McEnery eds. (2000), Kettemann and Marko eds. (2002), and Aston, Bernardini and Stewart eds. (2004).

corpora research and by other investigations of the indirect application type are extremely revealing and may help to improve pedagogical language descriptions.

Despite the research activities by the members of the TaLC community over the last ten years, a lot still remains to be done. Meunier as well as Granger, for instance, observe that "[l]earner corpus research is still in its infancy" (Meunier 2002, 123), and "few practical results have yet been achieved" in this particular field (Granger 2005: 131). I too would argue that there is certainly a need for more studies of the kind that have been mentioned in section 3.2 above, in particular for more comparative studies of a wider range of features in real English and 'school' English, maybe also covering different varieties of English (cf. Römer 2005, 296). I would suggest to complement such studies with learner corpus and contrastive analyses of problematic language features and to incorporate research findings in teaching materials. What we ought to take serious in this context is the criticism articulated by Flowerdew (1998: 550; as quoted in Granger 2004: 136), namely that in most studies of learner corpora "the implications for pedagogy are not developed in any great detail with the consequences that the findings have had little influence on [...] syllabus and materials design." Unfortunately, this is true not only for learner corpus studies but for corpus studies in general.

4.3 'Missionary work'

Most items on my CL & LT wish list fall into a category I have labelled 'missionary work'. The central idea here is to 'spread the word' among practitioners (i.e. teachers, students, materials writers, and syllabus designers) and to convince them that corpora can be of great use in their everyday work.

I would suggest starting these missionary activities at the level of initial teacher training at universities, and introduce students of English Studies (*Anglistik* or *Amerikanistik*), especially those who participate in an EFL teacher training programme, to corpus analysis at a relatively early stage in their course. We could even go as far as to make corpus linguistics an obligatory component of the university curriculum for future teachers of English. One important challenge will certainly be to get more students on our side, which may sound more straightforward than it is. In my experience, German students are often rather reluctant when it comes to accepting corpus evidence that runs counter to their intuition and to what they have learnt about the language at school. The corpus and the concordance seem to be the 'bad guys' who destroy their orderly world of clear-cut grammatical rules and clear right-or-wrong decisions. As Conrad (2004, 68) notes, "students seek only definite answers – such as being able to identify what is grammatical and ungrammatical" and tend to consider some characteristics of the language, such as variation, as "annoying aspect[s]". By introducing students to corpus analysis, we could show them that it is possible to identify regularities within variation and give them access to the fascinating patterned nature of language.

A next important step will be to reach teachers. Bernardini has recently observed that, despite researchers's efforts to promote DDL, "corpora seem not to have made

a splash in the classroom so far" (Bernadini 2005, 131). This indicates a need to convince practising teachers to use corpora and concordances in the classroom. Necessary prerequisites would of course be to provide them with a basic CL & LT training (for instance of the type sketched by Mukherjee 2004), and to equip them with computers, corpora, and concordance software. The task of the corpus linguist would be to show the teacher where corpora can help her/him solve everyday problems, e.g. when right-or-wrong decisions have to be made in exam marking. The problems secondary school teachers in Germany encounter when having to mark English A-level essays, for instance, are discussed in an interesting article by Dretzke (2001). He notes that, since most teachers are non-native speakers, some errors in the pupils' writing remain unnoticed while some correct passages are marked wrong (cf. Dretzke 2001, 44). Here, a corpus could function as an always available native-speaker consultant (cf. the quotation from Barnbrook in Section 3.1).

In addition to students and teachers, it would also be worth reaching syllabus designers and materials writers with the 'corpus mission'. Not only could materials writers draw on corpus research findings in making decisions about the presentation and sequencing of grammatical features in textbooks, a corpus could also help them to create exercises that are based on real language data and that present items in their most common collocations. For the syllabus designer, distributional and frequency information would be extremely valuable when planning a course and when solving problems of selection and progression. Last but not least, I think we should encourage advanced learners to get their hands on corpora themselves (maybe starting with what is freely available online) so they can explore the language and learn about its central patterns in an autonomous way.

5. Conclusion

In his introduction to the collection *How to Use Corpora in Language Teaching*, John Sinclair notes that "corpora are *almost* part of the pedagogical landscape" (Sinclair 2004b, 2; my emphasis). I hope that it has become clear from the discussion in the present paper that a lot still remains to be done before we can delete the "almost" from Sinclair's statement and before we can say that corpora have actually 'arrived' in language pedagogy.

I also hope to have shown that, despite the long list of desiderata, we have reason to be optimistic. Knowing about the unresolved issues and existing problems in the field and putting them on our CL & LT agenda may help to close the "large gap between the wealth of applied corpus-linguistic research and the teaching practice in Germany" that has been described by Mukherjee (2004, 247). My general recommendation would be to foster communication among and improve the exchange between researchers on the one side and practitioners (teachers, teacher trainers and trainees, and materials writers) on the other side, so that more people see that corpora are immensely valuable tools in a language learning and teaching context, and that it can be very profitable and, in fact, enjoyable for any teacher and learner to "enter the world of the corpus" (Sinclair 2004c, 297).

Works Cited

Aijmer, Karin (2002). "Modality in advanced Swedish learners' written interlanguage." Sylviane Granger, Joseph Hung and Stephanie Petch-Tyson, eds. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: Benjamins, 55-76.

Aston, Guy, Silvia Bernardini and Dominic Stewart, eds. (2004). *Corpora and Language Learners*. Amsterdam: John Benjamins.

Barlow, Michael (2000). "Parallel texts in language teaching." Simon Botley, Tony McEnery and Andrew Wilson, eds. *Multilingual Corpora in Teaching and Research*. Amsterdam: Rodopi, 106-115.

Barlow, Michael (2005). "Computer-based analyses of learner language." Gary Barkhuizen and Rod Ellis, eds. *Analysing Learner Language*. Oxford: Oxford University Press, 335-357.

Barnbrook, Geoff (1996). *Language and Computers. A Practical Introduction to the Computer Analysis of Language*. Edinburgh: Edinburgh University Press.

Bernardini, Silvia (2000). "Systematising serendipity: proposals for concordancing large corpora with language learners." Lou Burnard and Tony McEnery, eds. *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang, 225-234.

Bernardini, Silvia (2002). "Exploring new directions for discovery learning." Bernhard Kettemann and Georg Marko, eds. *Teaching and Learning by Doing Corpus Analysis. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000.* Amsterdam: Rodopi, 165-182.

Bernardini, Silvia (2005). "Corpora with language learners." *Abstracts Corpus Linguistics 2005*. Centre for Corpus Research, University of Birmingham, UK, 131.

Botley, Simon, Julia Glass, Tony McEnery and Andrew Wilson, eds. (1996). *Proceedings of Teaching and Language Corpora 1996.* Lancaster: University Centre for Computer Corpus Research on Language.

Breyer, Yvonne (2005). "Love's Labour's Lost: The troublesome relationship between corpus linguistics research and its application in EFL teacher training in Germany." Paper presented at the "Corpus Linguistics 2005" conference, 15-17 July 2005, University of Birmingham, UK.

Burnard, Lou and Tony McEnery, eds. (2000). *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang.

Claridge, Claudia (2002). "Translating phrasal verbs." Bernhard Kettemann and Georg Marko, eds. *Teaching and Learning by Doing Corpus Analysis. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000.* Amsterdam: Rodopi, 361-373.

Conrad, Susan (2004). "Corpus linguistics, language variation, and language teaching." John McH. Sinclair, ed. *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, 67-85.

Dretzke, Burkhard (2001). "Abiturkorrekturen Englisch oder die Last mit der Zweitkorrektur." *Der Fremdsprachliche Unterricht Englisch* 1, 44-46.

Flowerdew, Lynne (1998). "Corpus-linguistic techniques applied to textlinguistics." *System* 26, 541-552.

Gouverneur, Céline and Fanny Meunier (2004). "The treatment of phraseology in ELT textbooks." Paper presented at the '6th Teaching and Language Corpora conference (TaLC 6)', University of Granada, Spain, 6-9 July 2004.

Grabowski, Eva and Dieter Mindt (1995). "A corpus-based learning list of irregular verbs in English." *ICAME Journal* 19, 5-22.

Granger, Sylviane (1999). "Uses of tenses by advanced EFL learners: evidence from an error-tagged computer corpus." Hilde Hasselgård and Signe Oksefjell, eds. *Out of Corpora. Studies in Honour of Stig Johansson*. Amsterdam: Rodopi, 191-202.

Granger, Sylviane (2004). "Computer learner corpus research: current status and future prospects." Ulla Connor and Thomas A. Upton, eds. *Applied Corpus Linguistics. A Multi-Dimensional Perspective*. Amsterdam: Rodopi, 123-145.

Granger, Sylviane (2005). "Learner corpora in foreign language learning and teaching." *Abstracts Corpus Linguistics 2005*. Centre for Corpus Research, University of Birmingham, UK, 131.

Granger, Sylviane and Stephanie Tyson (1996). "Connector usage in the English essay writing of native and non-native EFL speakers of English." *World Englishes* 15, 19-29.

Hundt, Marianne, Nadja Nesselhauf and Carolin Biewer, eds. (forthcoming). *Corpus Linguistics and the Web.* Amsterdam: Rodopi.

Hunston, Susan (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Johns, Tim F. (1986). "Microconcord: A language-learner's research tool." *System* 14.2, 151-162.

Johns, Tim F. (1991). "Should you be persuaded – two samples of data-driven learning materials." Tim F. Johns and Philip King, eds. *Classroom Concordancing (ELR Journal* 4), 1-16.

Johns, Tim F. (2002). "Data-driven learning: The perpetual challenge." Bernhard Kettemann and Georg Marko, eds. *Teaching and Learning by Doing Corpus Analysis. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000.* Amsterdam: Rodopi, 107-117.

Kettemann, Bernhard and Georg Marko, eds. (2002). *Teaching and Learning by Doing Corpus Analysis. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000.* Amsterdam: Rodopi.

Meunier, Fanny (2002). "The pedagogical value of native and learner corpora in EFL grammar teaching." Sylviane Granger, Joseph Hung and Stephanie Petch-Tyson, eds. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: Benjamins, 119-141.

Meunier, Fanny (2004). "Native corpora, learner corpora and ELT: The winning team?" Paper presented at the '25th Conference of the International Computer Archive of Modern and Medieval English (ICAME 25)', University of Verona, Italy, 19-23 May 2004.

Mindt, Dieter (1981). "Linguistische Grammatik, didaktische Grammatik und pädagogische Grammatik." *Neusprachliche Mitteilungen* 34, 28-35.

Mindt, Dieter (1987). *Sprache – Grammatik – Unterrichtsgrammatik. Futurischer Zeitbezug im Englischen I*. Frankfurt: Diesterweg.

Mindt, Dieter (1997). "Corpora and the teaching of English in Germany." Anne Wichmann et al., eds. *Teaching and Language Corpora*. London: Longman, 40-50.

Mukherjee, Joybrato (2002). *Korpuslinguistik und Englischunterricht. Eine Einführung.* Frankfurt: Peter Lang.

Mukherjee, Joybrato (2004). "Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany." Ulla Connor and Thomas A. Upton, eds. *Applied Corpus Linguistics. A Multi-Dimensional Perspective*. Amsterdam: Rodopi, 239-250.

Nesselhauf, Nadja (2004). "Learner corpora and their potential for language teaching." John McH. Sinclair, ed. *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, 125-152.

Nesselhauf, Nadja (2005). *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.

Pravec, Norma A. (2002). "Survey of learner corpora." *ICAME Journal* 26, 81-114.

Römer, Ute (2004a). "A corpus-driven approach to modal auxiliaries and their didactics." John McH. Sinclair, ed. *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, 185-199.

Römer, Ute (2004b). "Comparing real and ideal language learner input: the use of an EFL textbook corpus in corpus linguistics and language teaching." Guy Aston, Silvia Bernardini and Dominic Stewart, eds. *Corpora and Language Learners*. Amsterdam: John Benjamins, 151-168.

Römer, Ute (2005). *Progressives, Patterns, Pedagogy. A Corpus-driven Approach to English Progressive Forms, Functions, Contexts and Didactics*. Amsterdam: John Benjamins.

Römer, Ute (forthcoming). "Where the computer meets language, literature, and pedagogy: Corpus analysis in English Studies." Andrea Gerbig and Anja Müller-Wood, eds. *Rethinking English: Reconciling Literature, Linguistics and Cultural Studies.* Lampeter: E. Mellen Press.

Schlüter, Norbert (2002). *Present Perfect. Eine korpuslinguistische Analyse des englischen Perfekts mit Vermittlungsvorschlägen für den Sprachunterricht*. Tübingen: Narr.

Schmied, Josef (1998). "Differences and similarities of close cognates: English *with* and German *mit*." S. Johansson and S. Oksefjell, eds. *Corpora and Cross-Linguistic Research. Theory, Method and Case Studies*. Amsterdam: Rodopi, 255-274.

Sinclair, John McH., ed. (1987). *Looking Up: An Account of the COBUILD Project in Lexical Computing.* London: Collins ELT.

Sinclair, John McH. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, John McH., ed. (2004a). *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins.

Sinclair, John McH. (2004b). "Introduction." *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, 1-10.

Sinclair, John McH. (2004c). "New evidence, new priorities, new attitudes." *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, 271-299.

Sinclair, John McH. and Malcolm Coulthard (1975). *Towards an Analysis of Discourse. The English Used by Teachers and Pupils.* London: Oxford University Press.

Wichmann, Anne, Steven Fligelstone, Tony McEnery and Gerry Knowles, eds. (1997). *Teaching and Language Corpora*. London: Longman.

Appendix 1: Relevant websites

American National Corpus (ANC) homepage:
http://americannationalcorpus.org/

AntConc concordance program by Laurence Anthony:
http://www.antlab.sci.waseda.ac.jp/antconc_index.html

BNC World online service for simple corpus searches:
http://thetis.bl.uk/lookup.html

British National Corpus (BNC) homepage:
http://www.natcorp.ox.ac.uk/

Collins Concordance and Collocations Sampler:
http://www.collins.co.uk/Corpus/CorpusSearch.aspx

ConcApp concordance program by Chris Greaves:
http://www.edict.com.hk/PUB/concapp/

International Computer Archive of Modern and Medieval English (ICAME):
http://helmer.aksis.uib.no/icame.html

International Corpus of English (ICE) project website:
http://www.ucl.ac.uk/english-usage/ice/

International Corpus of Learner English (ICLE) project website:
http://cecl.fltr.ucl.ac.be/Cecl-Projects/Icle/icle.htm

Michigan Corpus of Academic Spoken English (MICASE) online:
http://www.hti.umich.edu/m/micase/

MonoConc Pro information on the Athelstan website:
http://www.athel.com/mp.html

Passapong Sripicharn's DDL materials:
http://www.geocities.com/tonypgnews/units_index_pilot.htm

Santa Barbara Corpus of Spoken American English (SBCSAE) website:
http://www.ldc.upenn.edu/Projects/SBCSAE/

Tim Johns's DDL pages:
http://www.eisu.bham.ac.uk/johnstf/timconc.htm

VIEW (Variation In English Words and phrases) interface of the BNC:
http://view.byu.edu/

WebCorp – the 'Web as Corpus' search engine:
http://www.webcorp.org.uk

WordSmith Tools information on Mike Scott's homepage:
http://www.lexically.net/wordsmith/

Appendix 2: Corpus-based reference works and teaching materials

Abbs, Brian and Ingrid Freebairn, eds. (2005). *Longman Dictionary of Contemporary English*. Munich: Langenscheidt-Longman.

Barlow, Michael and Stephanie Burdine (forthcoming). *American Phrasal Verbs (CorpusLAB Series)*. Houston, TX: Athelstan.

Biber, Douglas, Geoffrey Leech and Susan Conrad (2002). *Longman Student Grammar of Spoken and Written English*. London: Longman.

Capel, Annette (1993). *Collins COBUILD Concordance Samplers 1: Prepositions*. London: HarperCollins.

Goodale, Malcolm (1995). *Collins COBUILD Concordance Samplers 4: Tenses*. London: HarperCollins.

Hill, Jimmie and Michael Lewis (1997). *LTP Dictionary of Selected Collocations*. Hove: Language Teaching Publications.

Hornby, A. S., ed. ([7]2005). *Oxford Advanced Learner's Dictionary*. Oxford: Oxford University Press.

Lea, Diana (2002). *Oxford Collocations Dictionary for Students of English*. Oxford: Oxford University Press.

McCarthy, Michael, Jeanne McCarten and Helen Sandiford (2005). *Touchstone Student's Book 1.* Cambridge: Cambridge University Press.

Mindt, Dieter (2000). *An Empirical Grammar of the English Verb*. Berlin: Cornelsen.

Peters, Pam (2004). *The Cambridge Guide to English Usage*. Cambridge: Cambridge University Press.

Rundell, Michael et al., eds. (2002). *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan.

Sinclair, John McH. (2003). *Reading Concordances*. London: Longman.

Sinclair, John McH. et al., eds. (1990). *Collins COBUILD English Grammar*. London: HarperCollins.

Sinclair, John McH. et al., eds. (1992). *Collins COBUILD English Usage*. London: HarperCollins.

Sinclair, John McH. et al., eds. ([3]2001). *Collins COBUILD English Dictionary for Advanced Learners*. London: HarperCollins.

Tribble, Chris and Glyn Jones (1997). *Concordances in the Classroom: A Resource Book for Teachers*. Houston, TX: Athelstan.