

7. Corpora and language teaching

1. Introduction
2. Indirect applications of corpora in language teaching
3. Direct applications of corpora in language teaching
4. Tasks for the future
5. Concluding remarks
6. Literature

1. Introduction

1.1. Corpus linguistics and language teaching

Over the past two decades, corpora (i. e. large systematic collections of written and/or spoken language stored on a computer and used in linguistic analysis) and corpus evidence have not only been used in linguistic research but also in the teaching and learning of languages – probably a use that “the compilers [of corpora] may not have foreseen” (Johansson 2007). There is now a wide range of fully corpus-based reference works (such as dictionaries and grammars) available to learners and teachers, and a number of dedicated researchers and teachers have made concrete suggestions on how concordances and corpus-derived exercises could be used in the language teaching classroom, thus significantly “[e]nriching the learning environment” (Aston 1997, 51). Indicative of the popularity of pedagogical corpora use and the need for research in this area is the considerable number of books and edited collections – some of which are the result of the successful “Teaching and Language Corpora” (TaLC) conference series – that have recently been published on the topic of this article or which bear a close relationship to it (cf. Ådel 2006; Aston 2001; Aston/Bernardini/Stewart 2004; Bernardini 2000a; Botley et al. 1996; Braun/Kohn/Mukherjee 2006; Burnard/McEney 2000; Connor/Upton 2004; Gavioli 2006; Ghadessy/Henry/Roseberry 2001; Granger/Hung/Petch-Tyson 2002; Hidalgo/Quereda/Santana 2007; Hunston 2002; Kettemann/Marko 2002; Mukherjee 2002; Nesselhauf 2005; Partington 1998; Römer 2005a; Schlüter 2002; Scott/Tribble 2006; Sinclair 2004a; Wichmann et al. 1997).

In this article I wish to examine the relationship between corpus linguistics (CL) and language teaching (LT) and provide an overview of the most important pedagogical applications of corpora. As Figure 7.1 aims to illustrate, this relationship is a dynamic one in which the two fields greatly influence each other. While LT profits from the resources, methods, and insights provided by CL, it also provides important impulses that are taken up in corpus linguistic research. The requirements of LT hence have an impact on research projects in CL and on the development of suitable resources and tools. The present article will investigate what influence CL has had on LT so far, and in what ways corpora have been used to improve pedagogical practice. It will also discuss further possible effects of CL on LT and of LT on CL, and highlight some future tasks for researchers and practitioners in the field.

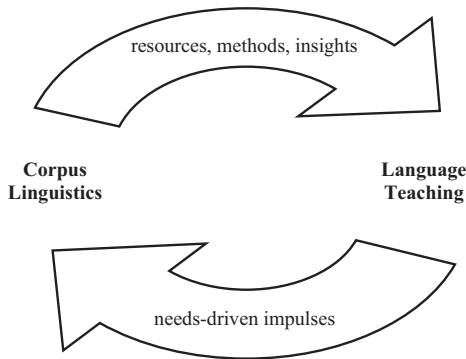


Fig. 7.1: The relationship between corpus linguistics (CL) and language teaching (LT)

1.2. Types of pedagogical corpus applications

When we talk about the application of corpora in language teaching, this includes both the use of corpus *tools*, i. e. the actual text collections and software packages for corpus access, and of corpus *methods*, i. e. the analytic techniques that are used when we work with corpus data. In classifying pedagogical corpus applications, i. e. the use of corpus tools and methods in a language teaching and language learning context, a useful distinction (going back to Leech 1997) can be made between direct and indirect applications. This means that, ‘indirectly’, corpora can help with decisions about what to teach and when to teach it, but that they can also be accessed ‘directly’ by learners and teachers in the LT classroom, and so “assist in the teaching process” (Fligelstone 1993, 98), thus affecting how something is taught and learnt. In addition to direct and indirect uses of

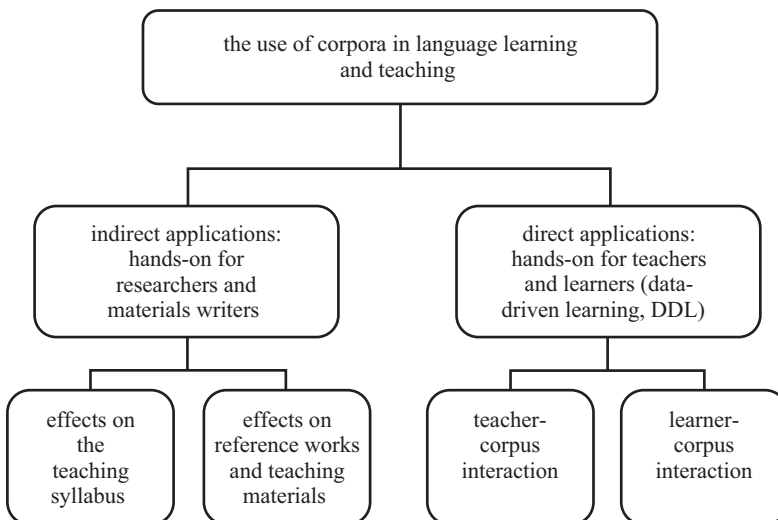


Fig. 7.2: Applications of corpora in language teaching

corpora in LT, Leech (1997, 5–6) talks about a third, in his opinion less-central component which he labels “further teaching-oriented corpus developments” (e. g. LSP corpora and learner corpora). These developments will, however, not be treated as marginal aspects here, but integrated in the discussion of direct and indirect pedagogical corpus applications. Sections 2 and 3 of this article will feature the most important lines of research and developments in both areas as presented in Figure 7.2. As the figure shows, we can identify different types of direct and indirect applications, depending on who or what is affected by the use of corpus methods and tools. In our discussions below, we will consider these distinctions and refer throughout to the pedagogical uses of general corpora, such as the BNC (British National Corpus, see articles 9, 10, 20), as well as specialised corpora, such as MICASE (the Michigan Corpus of Academic Spoken English, see articles 9, 47).

2. Indirect applications of corpora in language teaching

As Barlow (1996, 32) notes, “[t]he results of a corpus-based investigation can serve as a firm basis for both linguistic description and, on the applied side, as input for language learning.” This implies that corpora and the evidence derived from them can greatly affect course design and the content of teaching materials (see also Hunston 2002, 137). Existing pedagogical descriptions are evaluated in the light of “new evidence” (Sinclair 2004c, 271), and new decisions are made about the *selection* of language phenomena, the *progression* in the course, and the *presentation* of the selected items and structures (cf. Mindt 1981, 179; Römer 2005a, 287–291). This kind of indirect pedagogical corpus use benefits from research based on or driven by general and specialised corpora.

2.1. Indirect applications of general corpora

2.1.1. Corpora and the teaching syllabus

Large general corpora have proven to be an invaluable resource in the design of language teaching syllabi which emphasise communicative competence (cf. Hymes 1972, 1992) and which give prominence to those items that learners are most likely to encounter in real-life communicative situations. In the context of computer corpus-informed English language teaching syllabi, the first and probably most groundbreaking development was the design of the *Collins COBUILD English Course* (CCEC; Willis/Willis 1989), an offshoot of the pioneering COBUILD project in pedagogically oriented lexicography (cf. Sinclair 1987; articles 3 and 8). The contents of this new, corpus-driven “lexical syllabus” are “the commonest words and phrases in English and their meanings” (Willis 1990, 124). With its focus on lexis and lexical patterns, the CCEC responds to some of the most central findings of corpus research, namely that language is highly patterned in that it consists to an immense degree of repeated word-combinations, and that lexis and grammar are inseparably linked (cf. Hoey 2000; Hunston 2002; Hunston/Francis 2000; Partington 1998; Römer 2005a, 2005b; Sinclair 1991; Stubbs 1996; Tognini Bonelli 2001). Also worth mentioning is a much earlier attempt to improve further the teaching of

English vocabulary that was made long before the advent of computers and electronic corpora. In 1934, Michael West organised a conference “to discuss the part played by corpus-based word lists in the teaching of English as a foreign language” (Kennedy 1992, 327). About 20 years later, West’s (1953) *General Service List of English Words* (GSL) was published and has since then exerted great influence on curriculum design (cf. Kennedy 1992, 328; Willis 1990, 47). As the title indicates, West’s GSL suggests a syllabus that is based on words rather than on grammatical structures. It is also based on *frequently occurring* rather than on rare words. Of course, frequency of occurrence is not the only criterion that should influence decisions about the inclusion of items in the teaching syllabus (there are other relevant criteria, such as “range, availability, coverage and learnability (Mackey 1965, 188)” (Kennedy 1992, 340); cf. also Nation 1990, 21), but it is certainly an immensely important one (see also Aston 2000, 8; Leech 1997, 16). It can be safely assumed that learners will find it easier to develop both their receptive and productive skills when they are confronted with the most common lexical items of a language and the patterns and meanings with which they typically occur than when the language teaching input they get gives high priority to infrequent words and structures which the learners will only rarely encounter in real-life situations.

Another strand in applied corpus research that aims to inform the teaching syllabus and also stresses the importance of frequency of occurrence, examines language items in actual language use and compares the distributions and patterns found in general reference corpora (of speech and/or writing) with the presentations of the same items in teaching materials (coursebooks, grammars, usage handbooks). The starting point for these kinds of studies is usually language features that are known to cause perpetual problems to learners, for example, for German, discourse particles (Jones 1997), modal verbs (Jones 2000), the passive voice (Jones 2000) prepositions (Jones 1997), or, for English, future time expressions (Mindt 1987, 1997), *if*-clauses (Römer 2004b), irregular verbs (Grabowski/Mindt 1995), linking adverbials (Conrad 2004), modal verbs (Mindt 1995; Römer 2004a), the present perfect (Lorenz 2002; Schlüter 2002), progressive verb forms (Römer 2005a, 2006) and reflexives (Barlow 1996). For all these phenomena, researchers have found considerable mismatches between naturally-occurring German or English and the type of German or English that is put forward as a model in the examined teaching materials. They have, as a consequence, called for corpus-inspired adjustments in the language teaching syllabus (particularly as far as selection and progression are concerned) and for revised pedagogical descriptions which present a more adequate picture of the language as it is actually used. A case in point here is the misrepresentation of the functions and contextual patterns of English progressive forms in EFL teaching materials used in German schools. Progressives that refer to repeated actions or events, for example, are considerably more frequent in ‘real’ English than in textbook English where the common function “repeatedness” is rather neglected and the focus is on single continuous events (cf. Römer 2005a, 261–263).

2.2.2. Corpora and reference works and teaching materials

The results of the abovementioned corpus-coursebook comparisons do not only inform the language teaching curriculum but also help with decisions about the presentation of items and structures in reference works and teaching materials. Research on general

corpora has exerted a huge influence on reference publishing and has led to a new generation of dictionaries and grammar books. Nowadays, “people who have never heard of a corpus are using the products of corpus research.” (McEnery/Xiao/Tono 2006, 97) In the context of ELT (English Language Teaching), the publications in the Collins COBUILD series constitute a major achievement. Based on real English and compiled with the needs of the language learner in mind, the COBUILD dictionaries, grammars, usage guides, and concordance samplers (cf. Capel 1993; Carpenter 1993; Goodale 1995; Sinclair et al. 1990; Sinclair et al. 1992; Sinclair et al. 2001) offer teachers and learners more reliable information about the English language than any of the more traditional reference grammars or older non-corpus-based dictionaries. Two major advantages of the COBUILD and other corpus-based reference works for learners, e. g. those published in the past few years by Longman, Macmillan, OUP and CUP (cf. e. g. Abbs/Freebairn 2005; Biber/Leech/Conrad 2002; Hornby 2005; Peters 2004; Rundell et al. 2002) are that they incorporate corpus-derived findings on frequency distribution and register variation, and that they contain genuine instead of invented examples. Particularly worth mentioning here is the student version of the entirely corpus-based *Longman Grammar of Spoken and Written English* (Biber et al. 2002). The importance of presenting learners with authentic language examples has been stressed in a number of publications (cf. de Beaugrande 2001; Firth 1957; Fox 1987; Kennedy 1992; Römer 2004b, 2005a; Sinclair 1991, 1997). Kennedy (1992, 366), for instance, cautions that “invented examples can present a distorted version of typicality or an over-tidy picture of the system”, and Sinclair (1991, 5) calls it an “absurd notion that invented examples can actually represent the language better than real ones”. Thanks to the ‘corpus revolution’, the language learner can today choose from a range of reference works that are thoroughly corpus-based and that offer improved representations of the language she or he wants to study. While coursebooks and other materials used in the LT classroom have long been lagging behind this development and been rather unaffected by advances in CL (at least as far as the EFL (English as a Foreign Language) market is concerned), the first attempts are now being made to produce textbooks which draw on corpus research and are fully based on real-life data, i. e. on language that has in fact occurred (cf. Barlow/Burdine 2006; Carter/Hughes/McCarthy 2000; McCarthy/McCarten/Sandiford 2005).

Another branch of general corpora research that has exerted some influence on the design of reference works and, to a lesser extent, teaching materials is the area of phraseology and collocation studies. Scholars like Biber et al. (1999), Hunston/Francis (2000), Kjellmer (1984), Lewis (1993, 1997, 2000), Meunier/Gouverneur (2007), Nattinger (1980), Pawley/Syder (1983), and Sinclair/Renouf (1988) have emphasised the importance of recurring word combinations and prefabricated strings in a pedagogical context because of their great potential in fostering fluency, accuracy and idiomaticity. Although corpus-based collocation dictionaries (e. g. Hill/Lewis 1997; Lea 2002) are available, and although information on phraseology (i. e. about the combinations that individual words favour) is implicitly included in learners dictionaries in the word definitions and the selected corpus examples – and sometimes even explicitly described, e. g. in the grammar column in the COBUILD dictionaries and in the *COBUILD Grammar Patterns* reference books (cf. Francis/Hunston/Manning 1996, 1998) – such information and exercises on typical collocations are as yet largely missing from LT coursebooks (or they are inadequate; cf. Meunier/Gouverneur 2007). Like Hunston/Francis (2000, 272), I see a necessity in and “look forward to [more] information about patterns being incorporated in language teaching materials.”

2.2. Indirect applications of specialised corpora

Like general corpora, corpora of specialised texts (e. g. from one particular field of expertise, such as economics, or a narrowly defined group of speakers/writers, such as learners with a particular L1 and a certain level of proficiency) and research findings based on them can also be used to improve pedagogical practice and affect LT syllabi or the design of teaching materials. I would like to distinguish three different types of specialised corpora: LSP (language for special purposes) corpora, learner corpora, and parallel or translation corpora.

“LSP is the language that is used to discuss specialized fields of knowledge”, and it is the purpose of this language “to facilitate communication between people who wish to discuss a specialized subject” (Bowker/Pearson 2002, 25, 27). Corpora that capture a particular LSP, e. g. a corpus of Italian business letters or a corpus of English chemistry textbooks, can have a positive impact on the design of syllabi and materials of LSP courses. As Gavioli (2006, 23) states with reference to courses of English for special purposes (ESP), “working out basic items to be dealt with is a key teaching problem.” ESP corpora can help solve this problem. To give just two examples, Flowerdew (1993) demonstrates how frequency and concordance data from a corpus of English biology lectures and readings can be used in the creation of a course syllabus and teaching materials for students of science, and that such corpus-derived materials enable LSP teachers to teach those words and expressions (and those uses of them) that the learners will need later on in order to handle texts in their subject area. Focussing on academic English in general, Coxhead (2002) uses corpus evidence to compile an Academic Word List (AWL) which contains those vocabulary items that are most relevant and useful to the learners. Coxhead’s AWL has become an important tool in learning and teaching EAP (English for Academic Purposes). Other related studies deal with the pedagogical implications of corpora of English tourism industry texts (Lam 2007), meat technology English (Pereira de Oliveira 2003), or English letters of application (Henry/Roseberry 2001). Henry/Roseberry (2001, 121), for example, suggest that to compile genre-specific compendia or glossaries which they term “Language Pattern Dictionaries” based on their specialised corpus (see also Bowker/Pearson 2002, 137) would bring the learner more “success in job hunting” (Henry/Roseberry 2001, 117).

Studies on learner corpora, i. e. systematic computerised collections of the language produced by language learners (article 15), are also highly relevant for syllabus design (cf Aston 2000, 11; Granger 2002, 22) since they provide insights on “the needs of specific learner populations” (Meunier 2002, 125) and help to test teachers’ intuitions about whether a particular phenomenon is difficult or not (Granger 2002, 22). It has been shown how the findings of such studies (e. g. those based on the International Corpus of Learner English, “ICLE”, or on the German error-annotated learner corpus “Falko”, cf. Aijmer 2002; Altenberg/Granger 2001; Granger 1999; Leriko-Szymanska 2007; Lorenz 1999; Lüdeling et al. 2005; Nesselhauf 2004, 2005) can “enrich usage notes” in learners’ dictionaries (Granger 2002, 24), or how they “can provide useful insights into which collocational, pragmatic or discourse features should be addressed in materials design” (Flowerdew 2001, 376–377). Researchers like Granger (e. g. 2002, 2004) have also given the suggestion of linking up learner corpora work with contrastive analyses and using findings from corpora of the learner’s mother tongue to interpret the results of learner corpus studies. Contrastive work (i. e. research based on parallel or translation

corpora; article 16) is clearly invaluable for the selection of “elements the learner is likely to mistreat because they are different [...] from those in his [or her] native language” (Kjellmer 1992, 375). Parallel corpora are, in Teubert’s (2004, 188) words, “the repositories of source language units of meaning and their target language equivalents.” A corpus-enhanced knowledge of these equivalents (approached from different source language perspectives) is undoubtedly of use for language material developers and compilers of reference works (e. g. bilingual dictionaries), as is a knowledge about language items which cause translation problems for learners (cf. Schmied 1998 and article 54).

3. Direct applications of corpora in language teaching

While the indirect approach centres on the impact of corpus evidence on syllabus design or teaching materials, and is concerned with corpus access by researchers and – though to a lesser extent – materials designers, the direct approach is more teacher- and learner-focused. Instead of having to rely on the researcher as mediator and provider of corpus-based materials, language learners and teachers get their hands on corpora and concordancers themselves and find out about language patterning and the behaviour of words and phrases in an “autonomous” way (cf. Bernardini 2002, 165). Tim Johns, who, strongly supported by Tony Dudley-Evans and Philip King, pioneered direct corpus applications in grammar and vocabulary classes in the English for International Students Unit at the University of Birmingham in the 1980s (John Sinclair, personal communication), made the suggestion to “confront the learner as directly as possible with the data, and to make the learner a linguistic researcher” (Johns 2002, 108). Johns (1997, 101) also referred to the learner as a “language detective” and formulated the motto “Every student a Sherlock Holmes!” This method, in which there is either an interaction between the learner and the corpus or, in a more controlled way, between the teacher and the corpus (cf. Figure 7.2) is now widely known under the label “data-driven learning” or DDL (cf. Johns 1986, 1994). DDL activities with language learners can be based on (usually larger) general reference corpora or on (smaller) specialised corpora.

3.1. Direct applications of general corpora

Following Johns’ example, a number of researchers have discussed ways in which general corpora and concordances derived from them can be used by language learners. Bernardini (2002, 165), for instance, describes the positive effects of “corpus-aided discovery learning” with the BNC, and describes corpora as “rich sources of autonomous learning activities of a serendipitous kind” (ibid.; cf. also Bernardini 2000b, 2004). She sees the learner in the role of a “*traveller* instead of a *researcher*” (Bernardini 2000a, 131; italics in original), and is less “interested in the starting or end point of a learning experience” than in what the learner experiences in between, on her or his journey (Bernardini 2000a, 142). Kettemann (1995, 30) too stresses the exploratory aspect of DDL and considers concordancing in the ELT classroom “motivating and highly experiential” for the learner.

The DDL method of using learner-centred activities with the teacher as a facilitator of these activities has, however, not only been discussed with reference to English language teaching and English language corpora, but has also been applied in teaching other languages. Whistle (1999), for example, reports on introducing DDL activities to the teaching of French in order to supplement other CALL (computer-assisted language learning) tasks. Dodd (1997) and Jones (1997) show how corpora of written and spoken German can be exploited “to give students a richer language-learning experience in the foreign language environment” (Dodd 1997, 131), and Kennedy/Miceli (2001, 2002) suggest corpus consultation for learners of Italian. To give an example of a possible DDL task, learners could be asked to compile concordances of a pair of near-synonyms (such as ‘speak’ and ‘talk’ in English or ‘connaître’ and ‘savoir’ in French; cf. Chambers 2005, 117) and work out the differences in the collocational and phraseological behaviour of these words (see concordance samples in Figure 7.3). Further examples of DDL activities with English, German, Italian, and Spanish corpora are described in Aston (1997, 2001); Brodine (2001); Coffey (2007); Davies (2000, 2004); Dodd (1997); Fligelstone (1993); Gavioli (2001); Hadley (1997, 2001); Johns (1991, 2002); Stevens (1991); Sripicharn (2004); Tribble (1997); Zorzi (2001); and especially in Tribble/Jones (1997).

1 ery? Can I think. .. Well I 'd like to speak about the gallery I like to speak for m
2 d morning. Hello! Yeah, I 'd like to speak about the the squeeze on the benefits.
3 ught people might be less willing to speak at the meeting if they knew it was bein
4 d pointing. And er when you get up to speak at the conference, you have to give yo
5 re 's not quite as bad as when I had to speak for Amnesty on Radio Essex last year an
6 ke to speak about the gallery I like to speak for myself and er just the visual arts
7 't got a word . you 're not allowed to speak for the rest of the week. he 's hiding
8 Well football fans? well David I ca n't speak on behalf of Hibs, all I can say Mm. i
9 a member of the public I 'm not here to speak on behalf of the theatre at all. I migh
10 rhaps people could as actually come and speak to me afterwards, if they, if they thi
11 e will if you 'd like to come along and speak to him individually afterwards he will
12 hat, then. When do you start again? I speak to Stella now Do you speak, do you star
13 er appropriate way of doing it. Shall I speak to Paula about that then? Yeah. Paula,
14 er Okay. I 'll speak to Simon. I 'll speak to Simon erm about borrowing his P C at
15 ot to go home and do? Are you sure? Speak to me Yes Okay, right, one person from
16 rt is to go back one step, not just to speak to people who are experiencing hurt, bu
17 ikey. Speak up loud, you 've got ta be speak up loud and clear. No. Uniform. Unifor
18 I think that they may be frightened to speak up and that they 're scared that if the
19 of talking about. Excuse me could you speak up just a little bit? Yes yes er Thank
20 first floor. Oh I 'm sorry! Can you speak up then? Oh sorry! Have I Mm mm. Have

1 ney. All of that being said and Mr will talk a bit more about the figures when he com
2 have an M P here on the phone line and talk about this er proposed pay rise. They ca
3 ? Mm. . Erm I just want to go back and talk about a few things we just touched on ea
4 are? Erm Without drawing it, try and talk about a square. What would happen? Erm
5 Mm. Like we had you know, who could talk about experiences you know, in a, in a
6 ey include Gerry Addams and we we could talk about the other side, we could talk abo
7 ot , you know, I 'm not er gon na even talk about the disaster at airport things, yo
8 Mm. And basically, again can we just talk about what we 're trying to achieve? My
9 on, and Mr I know in a minute we 'll talk about the number of people who attend co
10 Becky. No, Becky. Becky . And we 'll talk about this in committee and let you kn
11 ople or all about that I do n't want to talk about that I want to talk I want to talk
12 get them around the table and begin to talk about those sort of things. Show a bit
13 bit, we just have done. We started to talk about the solar system. How far have we
14 them anyway. Right. What else did we talk about and we need to know, we have n't g
15 ber of possible pieces there, Did we talk about this? We did. Ah, right, yes. So
16 . They look tired and worn out. They talk quietly amongst themselves sometimes fin
17 s boss, the professor to come along and talk to us, and let, some of their time is
18 than then, because as a kid they do n't talk to you about them things do they like, y
19 going to talk to somebody who will not talk to you, who will not s possibly even sm
20 how nice the man was when he started to talk to Jason oh I 'll go and help you how fr

Fig. 7.3: Concordance samples of ‘speak’ and ‘talk’, based on the spoken part of the British National Corpus

Advantages of corpora work with learners have been formulated by scholars like Sinclair (1997, 38), who notes that, for the learner, “[c]orpora will clarify, give priorities, reduce exceptions and liberate the creative spirit.” Likewise, many researchers and teachers in

the TaLC (Teaching and Language Corpora) tradition are convinced that DDL can empower learners to find out things for themselves, and that corpora have a great pedagogic potential. The effectiveness of DDL has actually been proven in studies on the teaching and learning of vocabulary by Cobb (1997), Cresswell (2007) and Stevens (1991). Concordancing has not only been shown to be a useful way “to mimic the effects of natural contextual learning” (Cobb 1997, 314), researchers have recently also highlighted its use and usefulness for error correction in foreign or second language writing (cf. Bernardini 2004; Chambers 2005; Gaskell/Cobb 2004; Gray 2005). These studies demonstrate that corpora nicely complement existing reference works and that they may provide information which a dictionary or grammar book may not provide.

The immediate accessibility of authoritative information about the language is also a major advantage for language *teachers* who decide to interact with a corpus. As a recent survey on teachers’ needs has shown (cf. Römer forthcoming), teachers often require native-speaker advice on language points. Computer corpora that have been described as “tireless native-speaker informant[s], with rather greater potential knowledge of the language than the average native speaker” (Barnbrook 1996, 140), can offer help in such situations. In the sense of a modified type of DDL, teachers could also access corpora to create DDL exercises for learners, tailored to their learners’ proficiency level and their particular learning needs. Such exercises would enable teachers to “present the structures [they wish to introduce] and their lexis at the same time” (Francis/Sinclair 1994, 200). Emphasising the great potential of corpus analysis in a pedagogical context, Hunston/Francis (2000, 272) suggest that teachers, especially those in training, “should be encouraged to identify patterns as a grammar point for learners to notice”. Concordancing can certainly help teachers create a data-rich learning environment and “enrich their own knowledge of the language” (Barlow 1996, 30), as well as that of their pupils.

3.2. Direct applications of specialised corpora

Data-driven learning activities are not restricted to large general corpora but can also be based on the types of small and specialised corpora that have been discussed in section 2.2. of this article: LSP corpora, learner corpora and parallel corpora. Classroom concordancing in Johns’ or Bernardini’s sense is regarded as a useful tool in teaching LSP by a number of researchers and teachers in this field (for an overview, see Gavioli 2006, ch. 4). Mparutsa/Love/Morrison (1991) comment, for example, on the problems ESP students may have with general words (such as ‘price’) that are used in special ways and in particular (fixed) expressions in certain genres (cf. Brodine 2001, 157; cf. also Thurstun/Candlin 1998). In “[w]orking with corpora,” Gavioli (2006, 131) states, “ESP students become familiar with a productive idea of idiomatic language features, [and] they learn to use and adapt language patterns to their own needs”. In a similar vein, Bondi (2001, 159) discusses DDL in LSP contexts as a language awareness-raising strategy. Providing examples of useful worksheets, she points out that “[s]tudents of economics, for example, could become much better readers by developing an awareness of the forms and functions of different meta-argumentative expressions [e. g. ‘considers’ or ‘examined’] and by learning to understand the different role they play in the different genres”. Small and specialised corpora can also function as the source for DDL materials

in general language teaching (cf. Tribble 1997), e.g. in the teaching of conversational skills. This particular use is exemplified by Pérez Basanta and Rodríguez Martín (2007) who extract typical features of spoken English from a corpus of film transcripts (a collection of subtitles from movie DVDs) to be used in DDL tasks in EFL conversation classes.

What has been said above about the awareness-raising potential of DDL activities with LSP corpora is also true for DDL with learner corpora (article 15). Taking up Granger/Tribble's (1998) suggestion to combine data from native and learner corpora in the LT classroom, Meunier (2002, 130–134) presents examples and describes the advantages of DDL exercises with parallel native and learner concordances (cf. also Papp 2007). However, such exercises should, according to Meunier (2002, 134), “only be used, here and there, to complement native data and to illustrate [...] universally problematic areas” (e.g. verb or noun complementation). Seidlhofer (2000, 207) also comments on “using learner corpora for learning”, though with a shift in focus from learner errors to dealing with questions learners have about what they and their classmates have written. This focus on familiar texts (i. e. on texts the learners themselves have produced) ensures motivation and, in Seidlhofer's (2000, 222) terms, “the consideration of two equally crucial points of reference for learners: where they are, i. e. situated in their L2 learning contexts, and where they eventually (may) want to get to, i. e. close to the native-speaker language using capacity captured by L1 corpora.” Pedagogical applications of such “local learner corpora” (Seidlhofer 2002, 213) have also been discussed by Mukherjee/Rohrbach (2006; cf. also Turnbull/Burston 1998). In their paper on error analysis in a local learner corpus, the authors claim that their learners do “not only profit from the correction of their own mistakes but also from the analysis of their fellow-students' errors and their corrections” (Mukherjee/Rohrbach 2006, 225). Local learner corpora like the ones described by Seidlhofer and by Mukherjee/Rohrbach could easily be compiled by a larger number of teachers and lecturers by simply collecting their students' writings in electronic format, and subsequently serve as an exciting source of data to inspire the creation of DDL materials.

The third type of specialised corpus described in section 2.2., the parallel or translation corpus (i. e. a corpus that consists of original texts and their translations), also lends itself to the kind of DDL exploitation that we have envisaged for LSP and learner corpora. In coming to terms with the meaning(s) of an item in a foreign language, it can be extremely helpful for learners to create a parallel concordance and look at the translation equivalents of this item in their native language, or, the other way round, look for perhaps partly unknown translation equivalents of a selected native language item in the target language. Another promising use of parallel corpora in LT lies in highlighting collocational and phraseological differences between a word in the target language and its dictionary translation in the source language. Gavioli (1997), for example, reports that classroom work based on concordances of English ‘crucial’ and Italian ‘cruciale/cruciali’ has led her Italian students to illuminating findings about the different behaviour of these words. Johns (2002, 114) discusses the potential of parallel corpora for the creation of “reciprocal language materials”, i. e. “materials which could be used both to teach language A to speakers of language B, and language B to speakers of language A.” He provides examples of exercises derived from English-French parallel concordances. Similar exercises, but resulting from English-Chinese parallel concordancing, have been designed by Wang (2001; cf. also Ghadessy/Gao 2001). Parallel or comparable

corpora (i. e. text collections in different languages but of similar text types) have also been described as “aids in translation activities” (Zanettin 2001, 193) and as useful tools in the training of professional translators and interpreters (cf. e. g. Bernardini 2002; Bowker/Pearson 2002, ch. 11). Johansson (2007) gives examples of using the English-Norwegian Parallel Corpus (article 16) with a group of Norwegian students in solving the students’ learning problems and in dealing with problems of translation. Concordancing activities in the education of translators are described by Gavioli/Zanettin (1997) and by Bernardini (1997). According to Gavioli/Zanettin (1997, 6), comparable corpora provide “a repertoire of naturally occurring contexts in the target language onto which hypothesised translations can be mapped.” They hence “problematize the choices of the translator” (ibid.) or trainee translator, and help him or her find the most adequate and acceptable translation.

4. Tasks for the future

Despite the progress that has unquestionably been made in the field of pedagogical corpus applications, there is still scope for development. A number of tasks can be formulated to foster both the indirect and the direct use of corpora in language learning and teaching. Referring back to what has been mentioned in section 1.2., the next two sections will examine what further effects CL could have on LT and vice versa.

4.1. Fostering the indirect use of corpora in language teaching

We have seen that corpora and corpus evidence have already had an immense impact on teaching syllabi, teaching materials, and especially reference works like dictionaries or grammars (see the discussions in sections 2.1. and 2.2.). I would, however, argue that general and specialised corpora could be even better exploited to positively affect pedagogical practice. I am thinking here of further research activities that are inspired or driven by the needs of learners and language teaching practitioners. No matter how promising the advances made in the field of TaLC are, we still have a long way to go in providing more adequate descriptions of different types of language (different text types, registers and varieties), based on larger collections of data. I expect, for instance, that the place that LSP has in the language teaching domain will become increasingly significant in the future, and that more and better teaching materials tailored to the communication needs of students of economics or participants of business English courses, to mention only two groups of learners, will be required. These teaching and learning resources should ideally be based on expert performances (as opposed to apprentice performances; cf. Tribble 1997) in the selected field and, if possible, on large amounts of genuine language material. This implies the need to compile more and larger corpora of different types of written and especially spoken data. A task for the corpus researcher will then be to derive from these corpora those items and meanings that are most relevant for the learner group in question (cf. the publications in the *CorpusLab* series that are tailored to different groups of learners; e. g. Barlow/Burdine 2006). If we wish to tailor materials to learners’ needs and focus on language points that tend to be particu-

larly troublesome, we also ought to create more learner corpora of different kinds and find out more about the characteristics of learner language, so that, in the future, a larger number of dictionaries, grammars, and textbooks will not only be corpus- but also learner corpus-informed.

Further important insights to boost indirect corpus applications in LT could come from contrastive linguistic research on the basis of parallel or comparable corpora – another field of research in which significant developments are to be expected in the next few years (cf. articles 16 and 54). A number of comparative analyses of selected lexical-grammatical features in corpora and coursebooks have been carried out, mainly based on English language corpora and EFL teaching materials (cf. section 2.1.). More investigations of this type, in particular for different languages but also for different varieties of English, could help to isolate further mismatches between ‘real’ language and ‘school’ language, which could then lead to further improvements of teaching materials (cf. also Johansson/Stavestr and 1987, 147).

4.2. Fostering the direct use of corpora in language teaching

Although a lot is still left to be done as far as the indirect use of corpora in LT is concerned, there is probably even more scope for development with respect to direct applications. The gap between corpus linguistics and the teaching reality described by Mukherjee (2004), is still far too wide, and the extent to which corpora and concordances have actually been used in LT classrooms is, unfortunately, as yet fairly limited. Now that we know how beneficial corpus work can be to the learner, I think that it is the applied corpus linguist’s task to, as Chambers (2005) and Mukherjee (2004) call it, “popularise” corpus consultation and the work with corpus data in schools. In order to achieve this, some obstacles have of course to be overcome and a DDL-friendly environment has to be created. First of all, schools have to be equipped with corpus computers and appropriate software packages. For this purpose, new concordance programs that are appealing and easy to use may have to be written so that teachers and learners are not put off from working with corpora right away because the software is too complex or not user-friendly enough. John Sinclair (personal communication) has recently initiated a project which will provide broadband and corpus access for every classroom in Scotland by 2007 with the aim to support written literacy of the 12+-year-olds. We are thus coming closer to Fligelstone’s (1993, 100) hoped for scenario in which learners can access corpora whenever they want and simply “go to any of the labs, hit the icon which says ‘corpus’ and follow the instructions on the screen” – but we are not quite there yet. Projects like Sinclair’s Scotland project ought to be encouraged in different countries. An alternative to providing direct corpus access in the classroom would be to introduce learners and teachers to the resources that are accessible online and show them the potential of the Web as a huge resource of language data (article 18). Boulton/Wilhelm (2006) in this context talk about freely available corpus tools that learners have a right to use and that ought to be put in the hands of the learner.

A second and very important step towards creating a DDL-friendly environment will be to guide teachers and learners and give them a basic training in accessing corpora and in working with and evaluating concordances. Such a training is crucial because, as

Sinclair (2004b, 2) puts it, “a corpus is not a simple object, and it is just as easy to derive nonsensical conclusions from the evidence as insightful ones” (cf. also Gavioli 1997, 83). Guidance for teachers in how to read concordances and advice on what types of DDL exercises they could create can be found in Sinclair (2003) and Tribble/Jones (1997). Once they are familiar with the basics of corpus work and have learnt to deal with their new role as facilitators of autonomous learning activities, a follow-up task for teachers will be to “create conditions to make it [i. e. corpus work] relevant for” their learners (Gavioli 2006, 133) and to encourage DDL activities of an inductive and exploratory kind. It would probably also be helpful for learners and teachers of different languages if more DDL materials with ready-made exercises and photocopiable work-sheets on selected language points were available, but both groups might profit a lot more from getting their hands on corpora themselves.

5. Concluding remarks

This article has focussed on the relationship between corpus linguistics and language teaching. I hope to have shown that corpus resources and methods have a great potential to improve pedagogical practice and that corpora can be used in a number of ways, indirectly to inform teaching materials and reference works, or directly as language learning tools and repositories for the design of data-intensive teaching activities. I have also tried to make clear that a lot still remains to be done in research and practice before corpus linguistics will eventually ‘arrive’ in the classroom. Communication between corpus researchers and practitioners has to be improved considerably so that teachers and learners get the support they need and deserve.

As for the development of the CL-LT relationship and going back to what is shown in Figure 7.1 above, I would predict that the requirements of language learners and teachers will keep affecting corpus research and the creation of suitable tools and resources. In the future, more and more developments in corpus linguistics will probably be oriented towards language teaching and learning. Among other things I envisage a stronger emphasis on learner corpora, spoken language corpora, and specialised corpora – corpora that are tailored to the target learner group and its needs. As suitably formulated by Aston (2000, 16), “language pedagogy is increasingly designing its own corpora to its own criteria”. We do not know exactly how these criteria will develop in the next few decades. One thing that we can be sure of, however, is that the field of corpus linguistics and language teaching has an exciting future that both researchers and teachers can, and should, look forward to.

6. Literature

- Abbs, Brian/Freebairn, Ingrid (eds.) (2005), *Longman Dictionary of Contemporary English*. Munich: Langenscheidt-Longman.
- Ädel, Annelie (2006), *Metadiscourse in L1 an L2 English*. Amsterdam: John Benjamins.
- Aijmer, Karin (2002), Modality in Advanced Swedish Learners’ Written Interlanguage. In: Granger/Hung/Petch-Tyson 2002, 55–76.

- Altenberg, Bengt/Granger, Sylviane (2001), The Grammatical and Lexical Patterning of MAKE in Native and Non-native Student Writing. In: *Applied Linguistics* 22(2), 173–195.
- Aston, Guy (1997), Enriching the Learning Environment: Corpora in ELT. In: Wichmann et al. 1997, 51–64.
- Aston, Guy (2000), Corpora and Language Teaching. In: Burnard/McEnery 2000, 7–17.
- Aston, Guy (ed.) (2001), *Learning with Corpora*. Bologna: CLUEB and Houston, TX: Athelstan.
- Aston, Guy/Bernardini, Silvia/Stewart, Dominic (eds.) (2004), *Corpora and Language Learners*. Amsterdam: John Benjamins.
- Barlow, Michael/Burdine, Stephanie (2006), *American Phrasal Verbs*. (CorpusLAB Series.) Houston, TX: Athelstan. <http://www.corpuslab.com> (accessed 12 September 2006).
- Barlow, Michael (1996), Corpora for Theory and Practice. In: *International Journal of Corpus Linguistics* 1(1), 1–37.
- Barnbrook, Geoff (1996), *Language and Computers. A Practical Introduction to the Computer Analysis of Language*. Edinburgh: Edinburgh University Press.
- Beaugrande, Robert de (2001), *Twenty Challenges to Corpus Research. And how to Answer them*. <http://beaugrande.bizland.com/Twenty%20questions%20about%20corpus%20research.htm> (accessed 22 August 2006).
- Bernardini, Silvia (1997), A ‘Trainee’ Translator’s Perspective on Corpora. Paper presented at: *Corpus Use and Learning to Translate (CULT)*, Bertinoro, 14–15 November 1997. <http://www.sslmit.unibo.it/cultpaps/trainee.htm> (accessed through <http://web.archive.org>, 22 August 2006).
- Bernardini, Silvia (2000a), *Competence, Capacity, Corpora. A Study in Corpus-aided Language Learning*. Bologna: CLUEB.
- Bernardini, Silvia (2000b), Systematising Serendipity: Proposals for Concordancing Large Corpora with Language Learners. In: Burnard/McEnery 2000, 225–234.
- Bernardini, Silvia (2002), Exploring New Directions for Discovery Learning. In: Kettemann/Marko 2002, 165–182.
- Bernardini, Silvia (2004), Corpora in the Classroom: An Overview and Some Reflections on Future Developments. In: Sinclair 2004a, 15–36.
- Biber, Douglas/Johansson, Stig/Leech, Geoffrey/Conrad, Susan/Finegan, Edward (1999), *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Biber, Douglas/Leech, Geoffrey/Conrad, Susan (2002), *Longman Student Grammar of Spoken and Written English*. London: Longman.
- Bondi, Marina (2001), Small Corpora and Language Variation: Reflexivity across Genres. In: Ghadessy/Henry/Roseberry 2001, 135–174.
- Botley, Simon/Glass, Julia/McEnery, Tony/Wilson, Andrew (eds.) (1996), *Proceedings of Teaching and Language Corpora 1996*. Lancaster: University Centre for Computer Corpus Research on Language.
- Boulton, Alex/Wilhelm, Stephan (2006), Habeant Corpus – they Should Have the Body: Tools Learners Have the Right to Use. Paper presented at: *27e Congrès du GERAS “Cours et Corpus”*. Université de Bretagne Sud, Lorient, France, 23–25 March 2006.
- Bowker, Lynne/Pearson, Jennifer (2002), *Working with Specialized Language. A Practical Guide to Using Corpora*. London: Routledge.
- Braun, Sabine/Kohn, Kurt/Mukherjee, Joybrato (eds.) (2006), *Corpus Technology and Language Pedagogy*. Frankfurt: Peter Lang.
- Brodine, Ruey (2001), Integrating Corpus Work into a Academic Reading Course. In: Aston 2001, 138–176.
- Burnard, Lou/McEnery, Tony (eds.) (2000), *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang.
- Capel, Annette (1993), *Collins COBUILD Concordance Samplers 1: Prepositions*. London: HarperCollins.
- Carpenter, Edwin (1993), *Collins COBUILD English Guides 4: Confusable Words*. London: HarperCollins.

- Carter, Ronald/Hughes, Rebecca/McCarthy, Michael (2000), *Exploring Grammar in Context. Grammar Reference and Practice*. Cambridge: Cambridge University Press.
- Chambers, Angela (2005), Integrating Corpus Consultation in Language Studies. In: *Language Learning and Technology* 9(2), 111–125.
- Chambers, Angela (2005), Popularising Corpus Consultation by Language Learners and Teachers. In: Hidalgo/Quereda/Santana 2007, 3–16.
- Cobb, Tom (1997), Is there Any Measurable Learning from Hands-on Concordancing? In: *System* 25(3), 301–315.
- Coffey, Stephen (2007), Investigating Restricted Semantic Sets in a Large General Corpus: Learning Activities for Students of English as a Foreign Language. In: Hidalgo/Quereda/Santana 2007, 161–174.
- Connor, Ulla/Upton, Thomas A. (eds.) (2004), *Applied Corpus Linguistics. A Multi-dimensional Perspective*. Amsterdam: Rodopi.
- Conrad, Susan (2004), Corpus Linguistics, Language Variation, and Language Teaching. In: Sinclair 2004a, 67–85.
- Coxhead, Averil (2002), The Academic Word List: A Corpus-based Word List for Academic Purposes. In: Kettemann/Marko 2002, 73–89.
- Cresswell, Andy (2007), Getting to ‘Know’ Connectors? Evaluating Data-driven Learning in a Writing Skills Course. In: Hidalgo/Quereda/Santana 2007, 267–288.
- Davies, Mark (2000), Using Multi-million Word Corpora of Historical and Dialectal Spanish Texts to Teach Advanced Courses in Spanish Linguistics. In: Burnard/McEnery 2000, 173–185.
- Davies, Mark (2004), Student Use of Large, Annotated Corpora to Analyze Syntactic Variation. In: Aston et al. 2004, 259–269.
- Dodd, Bill (1997), Exploiting a Corpus of Written German for Advanced Language Learning. In: Wichmann et al. 1997, 131–145.
- Firth, John R. (1957), *Papers in Linguistics 1934–1951*. London: Oxford University Press.
- Fligelstone, Steve (1993), Some Reflections on the Question of Teaching, from a Corpus Linguistics Perspective. In: *ICAME Journal* 17, 97–109.
- Flowerdew, John (1993), Concordancing as a Tool in Course Design. In: *System* 21(2), 231–244 (reprinted in Ghadessy/Henry/Roseberry 2001, 71–92).
- Flowerdew, Lynne (2001), The Exploitation of Small Learner Corpora in EAP Materials Design. In: Ghadessy/Henry/Roseberry 2001, 363–380.
- Fox, Gwyneth (1987), The Case for Examples. In: Sinclair 1987, 137–149.
- Francis, Gill/Sinclair, John McH. (1994), ‘I Bet he Drinks Carling Black Label’: A Riposte to Owen on Corpus Grammar. In: *Applied Linguistics* 15(2), 190–202.
- Francis, Gill/Hunston, Susan/Manning, Elizabeth (1996), *Collins COBUILD Grammar Patterns 1: Verbs*. London: HarperCollins.
- Francis, Gill/Hunston, Susan/Manning, Elizabeth (1998), *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives*. London: HarperCollins.
- Gaskell, Delian/Cobb, Tom (2004), Can Learners Use Concordance Feedback for Writing Errors? In: *System* 32, 301–319.
- Gavioli, Laura/Zanettin, Federico (1997), Comparable Corpora and Translation: A Pedagogic Perspective. Paper presented at: *Corpus Use and Learning to Translate (CULT)*. Bertinoro, 14–15 November 1997. <http://www.sslmit.unibo.it/cultpaps/laura-fede.htm> (accessed through <http://web.archive.org>, 22 August 2006).
- Gavioli, Laura (1997), Exploring Texts through the Concordancer: Guiding the Learner. In: Wichmann et al. 1997, 83–99.
- Gavioli, Laura (2001), The Learner as Researcher: Introducing Corpus Concordancing in the Classroom. In: Aston 2001, 108–137.
- Gavioli, Laura (2006), *Exploring Corpora for ESP Learning*. Amsterdam: John Benjamins.
- Ghadessy, Mohsen/Gao, Yanjie (2001), Small Corpora and Translation: Comparing Thematic Organization in Two Languages. In: Ghadessy/Henry/Roseberry 2001, 335–359.

- Ghadessy, Mohsen/Henry, Alex/Roseberry, Robert L. (eds.) (2001), *Small Corpus Studies and ELT. Theory and Practice*. Amsterdam: John Benjamins.
- Goodale, Malcolm (1995), *Collins COBUILD Concordance Samplers 4: Tenses*. London: HarperCollins.
- Grabowski, Eva/Mindt, Dieter (1995), A Corpus-based Learning List of Irregular Verbs in English. In: *ICAME Journal* 19, 5–22.
- Granger, Sylviane (1999), Uses of Tenses by Advanced EFL Learners: Evidence from an Error-tagged Computer Corpus. In: Hasselgård/Oksefjell 1999, 191–202.
- Granger, Sylviane (2002), A Bird's-eye View of Learner Corpus Research. In: Granger/Hung/Petch-Tyson 2002, 3–33.
- Granger, Sylviane (2004), Computer Learner Corpus Research: Current Status and Future Prospects. In: Connor/Upton 2004, 123–145.
- Granger, Sylviane/Hung, Joseph/Petch-Tyson, Stephanie (eds.) (2002), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.
- Granger, Sylviane/Tribble, Chris (1988), Learner Corpus Data in the Foreign Language Classroom: Form-focused Instruction and Data-driven Learning. In: Granger, Sylviane (ed.), *Learner English on Computer*. London: Longman, 199–209.
- Gray, Bethany E. (2005), *Error-specific Concordancing for Intermediate ESL/EFL Writers*. <http://dana.ucc.na4.edu/~bde6/coursework/Projects/PedagogicalTip/homepage.html> (accessed 28 January 2008).
- Hadley, Gregory (1997), *Sensing the Winds of Change: An Introduction to Data-driven Learning*. <http://www.nuis.ac.jp/~hadley/publication/windofchange/windofchange.htm> (accessed 22 August 2006).
- Hadley, Gregory (2001), *Concordancing in Japanese TEFL: Unlocking the Power of Data-driven Learning*. <http://www.nuis.ac.jp/~hadley/publication/jlearner/jlearner.htm> (accessed 22 August 2006).
- Henry, Alex/Roseberry, Robert L. (2001), Using a Small Corpus to Obtain Data for Teaching a Genre. In: Ghadessy/Henry/Roseberry 2001, 93–133.
- Hidalgo, Encarnacion/Quereda, Luis/Santana, Juan (eds.) (2007), *Corpora in the Foreign Language Classroom*. Amsterdam: Rodopi.
- Hill, Jimmie/Lewis, Michael (1997), *LTP Dictionary of Selected Collocations*. Hove: Language Teaching Publications.
- Hoey, Michael (2000), The Hidden Lexical Clues of Textual Organisation: A Preliminary Investigation into an Unusual Text from a Corpus Perspective. In: Burnard/McEnery 2000, 31–41.
- Hornby, Albert S. (ed.) (2005), *Oxford Advanced Learner's Dictionary*. Oxford: Oxford University Press.
- Hunston, Susan/Francis, Gill (2000), *Pattern Grammar. A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- Hunston, Susan (2002), *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hymes, Dell (1972), On Communicative Competence. In: Pride, John B./Holmes, Janet (eds.), *Sociolinguistics*. Harmondsworth: Penguin, 269–293.
- Hymes, Dell (1992), The Concept of Communicative Competence Revisited. In: Pütz, Martin (ed.), *Thirty Years of Linguistic Evolution. Studies in Honour of Rene Dirven on the Occasion of his Sixtieth Birthday*. Amsterdam: John Benjamins, 31–57.
- Johansson, Stig (2007), Using Corpora: From Learning to Research. In: Hidalgo/Quereda/Santana 2007, 17–30.
- Johansson, Stig/Stavestrand, Helga (1987), Problems in Learning – and Teaching – the Progressive Form. In: Lindblad, Ishrat/Ljung, Magnus (eds.), *Proceedings from the Third Nordic Conference for English Studies, Hässelby, Sept 25–27, 1986, vol. 1*. Stockholm: Almqvist & Wiksell, 139–148.
- Johns, Tim (1986), Microconcord: A Language-learner's Research Tool. In: *System* 14(2), 151–162.

- Johns, Tim (1991), Should you Be Persuaded – Two Samples of Data-driven Learning Materials. In: Johns, Tim/King, Philip (eds.), *Classroom Concordancing (ELR Journal 4)*, 1–16.
- Johns, Tim (1994), From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-driven Learning. In: Odlin, Terence (ed.), *Perspectives on Pedagogical Grammar*. Cambridge: Cambridge University Press, 27–45.
- Johns, Tim (1997), Contexts: The Background, Development and Trialling of a Concordance-based CALL Program. In: Wichmann et al. 1997, 100–115.
- Johns, Tim (2002), Data-driven Learning: The Perpetual Challenge. In: Kettemann/Marko 2002, 107–117.
- Jones, Randall L. (1997), Creating and Using a Corpus of Spoken German. In: Wichmann et al. 1997, 146–156.
- Jones, Randall L. (2000), Textbook German and Authentic Spoken German: A Corpus-based Comparison. In: Lewandowska-Tomaszczyk, Barbara/Melia, Patrick James (eds.), *PALC'99: Practical Applications in Language Corpora*. Frankfurt: Peter Lang, 501–516.
- Kennedy, Claire/Miceli, Tiziana (2001), An Evaluation of Intermediate Students' Approaches to Corpus Investigation. In: *Language Learning and Technology 5*(3), 77–90. <http://llt.msu.edu/vol5num3/kennedy/> (accessed 22 August 2006).
- Kennedy, Claire/Miceli, Tiziana (2002), The CWIC Project: Developing and Using a Corpus for Intermediate Italian Students. In: Kettemann/Marko 2002, 183–192.
- Kennedy, Graeme (1992), Preferred Ways of Putting Things with Implications for Language Teaching. In: Svartvik 1992, 335–373.
- Kettemann, Bernhard/Marko, Georg (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.
- Kettemann, Bernhard (1995), On the Use of Concordancing in ELT. In: *Arbeiten aus Anglistik und Amerikanistik 20*, 29–41.
- Kjellmer, Göran (1984), Some Thoughts on Collocational Distinctiveness. In: Aarts, Jan/Meijs, Willem (eds.), *Corpus Linguistics. Recent Developments in the Use of Computer Corpora in English Language Research*. Amsterdam: Rodopi, 163–171.
- Kjellmer, Göran (1992), Comments. In: Svartvik 1992, 374–378.
- Lam, Peter Y. W. (2007), A Corpus-driven Lexico-grammatical Analysis of English Tourism Industry Texts and the Study of its Pedagogic Implications in English for Specific Purposes. In: Hidalgo/Quereda/Santana 2007, 71–90.
- Lea, Diana (ed.) (2002), *Oxford Collocations Dictionary for Students of English*. Oxford: Oxford University Press.
- Leech, Geoffrey (1997), Teaching and Language Corpora: A Convergence. In: Wichmann et al. 1997, 1–23.
- Leńko-Szymańska, Agnieszka (2007), Past Progressive or Simple Past? The Acquisition of Progressive Aspect by Polish Advanced Learners of English. In: Hidalgo/Quereda/Santana 2007, 253–266.
- Lewis, Michael (1993), *The Lexical Approach*. Hove: Language Teaching Publications.
- Lewis, Michael (1997), *Implementing the Lexical Approach*. Hove: Language Teaching Publications.
- Lewis, Michael (2000), *Teaching Collocation. Further Developments in the Lexical Approach*. Hove: Language Teaching Publications.
- Lorenz, Gunter (1999), *Adjective Intensification – Learners versus Native Speakers. A Corpus Study of Argumentative Writing*. Amsterdam: Rodopi.
- Lorenz, Gunter (2002), Language Corpora Rock the Base: On Standard English Grammar, Perfective Aspect and Seemingly Adverse Corpus Evidence. In: Kettemann/Marko 2002, 131–145.
- Lüdeling, Anke/Walter, Maik/Kroymann, Emil/Adolphs, Peter (2005), Multi-level Error Annotation in Learner Corpora. In: Hunston, Susan/Danielsson, Pernilla (eds.), *Proceedings from the Corpus Linguistics Conference Series (Corpus Linguistics 2005, Birmingham, 14–15 July 2005)*. <http://www.corpus.bham.ac.uk/PCLC> (accessed 22 August 2006).

- Mackey, William Francis (1965) *Language teaching analysis*. London: Longman.
- McCarthy, Michael/Jeanne McCarten/Sandiford, Helen (2005), *Touchstone Student's Book 1*. Cambridge: Cambridge University Press.
- McEnery, Tony/Xiao, Richard/Tono, Yukio (2006), *Corpus-based Language Studies. An Advanced Resource Book*. London: Routledge.
- Meunier, Fanny/Gouverneur, Celine (2007), The Treatment of Phraseology in ELT Textbooks. In: Hidalgo/Quereda/Santana 2007, 119–140.
- Meunier, Fanny (2002), The Pedagogical Value of Native and Learner Corpora in EFL Grammar Teaching. In: Granger/Hung/Petch-Tyson 2002, 119–141.
- Mindt, Dieter (1981), Angewandte Linguistik und Grammatik für den Englischunterricht. In: Kuns- mann, Peter/Kuhn, Otto (eds.), *Weltsprache Englisch in Forschung und Lehre. Festschrift für Kurt Wächtler*. Berlin: Schmidt, 175–186.
- Mindt, Dieter (1987), *Sprache – Grammatik – Unterrichtsgrammatik. Futurischer Zeitbezug im Englischen I*. Frankfurt: Diesterweg.
- Mindt, Dieter (1995), *An Empirical Grammar of the English Verb: Modal Verbs*. Berlin: Cornelsen.
- Mindt, Dieter (1997), Corpora and the Teaching of English in Germany. In: Wichmann et al. 1997, 40–50.
- Mparutsa, Cynthia/Love, Alison/Morrison, Andrew (1991), Bringing Concord to the ESP Classroom. In: Johns, Tim/King, Philip (eds.), *Classroom Concordancing (ELR Journal 4)*, 115–134.
- 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. In: Connor/Upton 2004, 239–250.
- Mukherjee, Joybrato/Rohrbach, Jan-Marc (2006), Rethinking Applied Corpus Linguistics from a Language-pedagogical Perspective: New Departures in Learner Corpus Research. In: Kette- mann, Bernhard/Marko, Georg (eds.), *Planing, Gluing and Painting Corpora. Inside the Applied Corpus Linguist's Workshop*. Frankfurt: Peter Lang, 205–232.
- Nation, Paul (1990), *Teaching and Learning Vocabulary*. Boston: Heinle & Heinle.
- Nattinger, James R. (1980), A Lexical Phrase Grammar for ESL. In: *TESOL Quarterly* 14(3), 337–344.
- Nesselhauf, Nadja (2004), How Learner Corpus Analysis Can Contribute to Language Teaching: A Study of Support Verb Constructions. In: Aston/Bernardini/Stewart 2004, 109–124.
- Nesselhauf, Nadja (2005), *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.
- Papp, Szilvia (2007), Inductive Learning and Self-correction with the Use of Learner and Reference Corpora. In: Hidalgo/Quereda/Santana 2007, 207–220.
- Partington, Alan (1998), *Patterns and Meanings. Using Corpora for English Language Research and Teaching*. Amsterdam: John Benjamins.
- Pawley, Andrew/Syder, Frances H. (1983), Two Puzzles for Linguistic Theory: Native-like Selection and Native-like Fluency. In: Richards, Jack C./Schmidt, Richard W. (eds.), *Language and Communication*. London: Longman, 191–226.
- Pereira de Oliveira, Maria José (2003), Corpus Linguistics in the Teaching of ESP and Literary Studies. In: *ESP World* 6(2). http://www.esp-world.info/articles_6/Corpus.htm (accessed 22 August 2006).
- Pérez Basanta, Carmen/Martín, María Elena Rodríguez (2007), The Application of Data-driven Learning to a Small-scale Corpus: Using Film Transcripts for Teaching Conversational Skills. In: Hidalgo/Quereda/Santana 2007.
- Peters, Pam (2004), *The Cambridge Guide to English Usage*. Cambridge: Cambridge University Press.
- Römer, Ute (2004a), A Corpus-driven Approach to Modal Auxiliaries and their Didactics. In: Sinclair 2004a, 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. In: Aston/Bernardini/Stewart 2004, 151–168.

- Römer, Ute (2005a), *Progressives, Patterns, Pedagogy. A Corpus-driven Approach to English Progressive Forms, Functions, Contexts and Didactics*. Amsterdam: John Benjamins.
- Römer, Ute (2005b), Shifting Foci in Language Description and Instruction. In: *Arbeiten aus Anglistik und Amerikanistik* 30(1–2), 145–160.
- Römer, Ute (2006), Looking at *Looking*: Functions and Contexts of Progressives in Spoken English and 'School' English. In: Renouf, Antoinette/Kehoe, Andrew (eds.), *The Changing Face of Corpus Linguistics*. Amsterdam: Rodopi, 231–242.
- Römer, Ute (forthcoming), Corpus Research and Practice: What Help Do Teachers Need and what Can we Offer? In: Aijmer, Karin (ed.), *Teaching and Corpora* (provisional title). Amsterdam: John Benjamins.
- Rundell, Michael (ed.) (2002), *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan.
- 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*. In: Johansson, Stig/Oksefjell, Signe (eds.), *Corpora and Cross-linguistic Research. Theory, Method and Case Studies*. Amsterdam: Rodopi, 255–274.
- Scott, Mike/Tribble, Christopher (2006), *Textual Patterns. Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins.
- Seidlhofer, Barbara (2000), Operationalizing Intertextuality: Using Learner Corpora for Learning. In: Burnard/McEnery 2000, 207–223.
- Seidlhofer, Barbara (2002), Pedagogy and Local Learner Corpora: Working with Learning-driven Data. In: Granger/Hung/Petch-Tyson 2002, 213–234.
- Sinclair, John McH. (ed.) (1987), *Looking Up. An Account of the COBUILD Project in Lexical Computing*. London: HarperCollins.
- Sinclair, John McH. (1991), *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Sinclair, John McH. (1997), Corpus Evidence in Language Description. In: Wichmann et al. 1997, 27–39.
- Sinclair, John McH. (2003), *Reading Concordances*. London: Longman.
- Sinclair, John McH. (ed.) (2004a), *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins.
- Sinclair, John McH. (2004b), Introduction. In: Sinclair 2004a, 1–10.
- Sinclair, John McH. (2004c), New Evidence, New Priorities, New Attitudes. In: Sinclair 2004a, 271–299.
- Sinclair, John McH./Renouf, Antoinette (1988), A Lexical Syllabus for Language Learning. In: Carter, Ronald/McCarthy, Michael (eds.), *Vocabulary in Language Teaching*. London: Longman, 140–158.
- 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.) (2001), *Collins COBUILD English Dictionary for Advanced Learners*. London: HarperCollins.
- Sripicharn, Passapong (2004), Examining Native Speakers' and Learners' Investigation of the Same Concordance Data and its Implications for Classroom Concordancing with EFL Learners. In: Aston/Bernardini/Stewart 2004, 233–245.
- Stevens, Vance (1991), Concordance-based Vocabulary Exercises: A Viable Alternative to Gap-fillers. In: Johns, Tim/King, Philip (eds.), *Classroom Concordancing (ELR Journal 4)*, 47–63.
- Stubbs, Michael (1996), *Text and Corpus Analysis: Computer-assisted studies of language and culture*. Oxford: Blackwell.
- Svartvik, J. (ed.) (1992), *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82 Stockholm, 4–8 August 1991*. Berlin/New York: Mouton de Gruyter.

- Teubert, Wolfgang (2004), Units of Meaning, Parallel Corpora, and their Implications for Language Teaching. In: Connor/Upton 2004, 171–189.
- Thurstun, Jennifer/Candlin, Christopher (1998), Concordancing and the Teaching of the Vocabulary of Academic English. In: *English for Specific Purposes* 17, 267–280.
- Tognini Bonelli, Elena (2001), *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Tribble, Chris (1997), Improvising Corpora for ELT: Quick-and-dirty Ways of Developing Corpora for Language Teaching. In: Lewandowska-Tomaszczyk, Barbara/Melia, Patrick J. (eds.), *Practical Applications in Language Corpora. The Proceedings of PALC '97*. Lodz: Lodz University Press. <http://www.ctribble.co.uk/text/Palc.htm> (accessed 22 August 2006).
- Tribble, Chris/Jones, Glyn (1997), *Concordances in the Classroom. A Resource Book for Teachers*. Houston, TX: Athelstan.
- Turnbull, Jill/Burston, Jack (1998), Towards Independent Concordance Work for Students: Lessons from a Case Study. In: *On-CALL* 12(2). <http://www.cltr.uq.edu.au/oncall/turnbull122.html> (accessed through <http://web.archive.org>, 22 August 2006).
- Wang, Lixun (2001), Exploring Parallel Concordancing in English and Chinese. In: *Language Learning and Technology* 5(3), 174–184.
- West, Michael (1953), *A General Service List of English Words*. London: Longman.
- Whistle, Jeremy (1999), Concordancing with Students Using an ‘off-the-Web’ Corpus. In: *ReCALL* 11(2), 74–80. Available at: <http://www.eurocall-languages.org/recall/pdf/rvol11no2.pdf>.
- Wichmann, Anne/Fligelstone, Steven/McEnery, Tony/Knowles, Gerry (eds.) (1997), *Teaching and Language Corpora*. London: Longman.
- Willis, Dave/Willis, Jane (1989), *Collins COBUILD English Course*. London: HarperCollins.
- Willis, Dave (1990), *The Lexical Syllabus. A New Approach to Language Teaching*. London: HarperCollins.
- Zanettin, Federico (2001), Swimming in Words: Corpora, Translation, and Language Learning. In: Aston 2001, 177–197.
- Zorzi, Daniela (2001), The Pedagogic Use of Spoken Corpora. In: Aston 2001, 85–107.

Ute Römer, Hannover (Germany)

8. Corpus linguistics and lexicography

1. Introduction
2. Corpora as a source for lexicography: Corpus composition
3. Corpora as a source for lexicography: Using corpus data as evidence
4. Corpus linguistic tools for lexicography
5. Two-way interaction between lexicography and corpus linguistics
6. Historical note
7. Literature

1. Introduction

1.1. Objectives and structure of this article

The production of dictionaries is one of the “clients” of corpus linguistics, insofar as many dictionaries of recent date have been created in some way “on the basis” of corpora. This reliance on corpus data concerns the selection of raw material from which