

Bridging the Gap between Applied Corpus Linguistics and the Reality of English Language Teaching in Germany

Joybrato Mukherjee

University of Giessen

Abstract

The starting point for the present paper is the results of a survey among English language teachers in German secondary schools. The survey shows that the practice of English language teaching in Germany is still largely unaffected by descriptive corpus-linguistic research into authentic language use and applied corpus-linguistic suggestions of using corpus resources and corpus-based methods for teaching purposes. In the light of this gap between applied corpus linguistics and the reality of English language teaching in Germany, it is suggested that a concerted effort is needed to popularise the language-pedagogical potential of corpus linguistics, preferably under the auspices of the local state teaching boards. In this context, particular attention should be paid to the preconceptions and needs of the vast majority of teachers who, for a variety of reasons, have not yet worked with corpora. In particular, it is necessary to implement teacher-centred corpus activities in the classroom before truly learner-centred methods are envisaged.

1 Introduction

Corpus linguists have shown a persistent interest in the language-pedagogical implications and applications of corpus-based research for several decades. The COBUILD project, resulting in a new generation of learner dictionaries (see Sinclair 1987), the early coinage of the notion of “data-driven learning” (see Johns 1991) and the compilation and analysis of learner corpora such as ICLE (see Granger 1998) provide ample testimony of this fact. At first blush, then, one might readily expect that the multitude of suggestions on how to use corpus data, corpus-based resources and corpus-linguistic methods in the English language classroom (see Burnard and McEnery 2000; Aston 2001; Mukherjee 2002) has already revolutionised – or is just about to do so – the way in which English is taught and learned as a foreign language. However, in Germany (and probably in many other countries as well) this turns out to be wishful thinking. In reality, the influence of applied corpus-linguistic research on the actual practice of English language teaching is still relatively limited. Tribble (2000: 31), for example, admits that “not many teachers seem to be *using* corpora in their classrooms.”

In order to empirically assess the extent to which English language teachers in Germany make use of – and actually know about – corpora, I conducted a survey in which 248 qualified English language teachers at secondary schools in North Rhine-Westphalia, the by far most heavily populated federal state of Germany, took part.¹ The survey data were collected in the

context of several advanced teacher training workshops on corpus linguistics for qualified English language teachers that took place in 2001 and 2002.

The idea to conduct such *test workshops* arose out of the desire to: firstly, introduce teachers of English at secondary schools to basic principles and methods in corpus linguistics; secondly, familiarize them with language-pedagogical applications and implications of corpus-based research; thirdly, find out what they know about corpus linguistics before the test workshop and what they think about the relevance of corpus linguistics to their own classroom practice after the test workshop. I am using the term *test workshop* in this context because at this stage the workshops were offered and carried out on an *ad hoc* basis whenever particular schools were willing to host such workshops for their English teachers as voluntary participants. In total, eight half-day test workshops took place. They were designed in slightly different ways in order to find out which of the formats would be most appropriate for an institutionalized introductory workshop to be offered to interested teachers by the local state teaching board (see section 3). What they had in common was the overall structure:

- In a lecture of about one hour, the participants were provided with a general introduction to some key concepts in corpus linguistics (e.g., major corpora of present-day English, the notion of representativeness, word-lists and concordances).
- In a seminar of about an hour, the participants were provided with selected findings from corpus-based research (e.g., concordances) that they had to compare with the descriptive statements given in traditional school textbooks and learner grammars. Of course, the examples focused on those fields in which there is a clear discrepancy between corpus data and traditional learner grammars (e.g., with regard to the use of *some* and *any*) or in which corpus data would give access to data not available otherwise (e.g., frequent lexicogrammatical patterns of a given word).
- In a practical part of about two hours, the participants were introduced to some applications of corpus data in the classroom that have been discussed in applied corpus linguistics (e.g., the production of concordance-based exercises). Also, some problems of usage were discussed in the light of corpus data (e.g., the question as to whether *example for*, *typical for* etc. – instead of *example of*, *typical of* etc. – occur in native usage or not).

All participants in the test workshops were asked to fill in a questionnaire; some questions were asked before the workshop started, others at the end of the workshop. It is the result of this survey to which I will turn in the following section. The participants' answers reveal that corpus-based methods have not yet exerted much influence on teaching practice in the English classroom in Germany. After discussing the survey results (see section 2), I will sketch out how corpus linguistics may be popularised in the German context (see

section 3), which may best be achieved by taking into consideration and focusing on the average teacher's preconceptions and needs (see section 4). Finally, I will offer a few concluding remarks on the implications of the survey data and the experiences from the test workshops.

Table 1: The role of corpus linguistics in English language teaching in Germany: some survey data

1) Before the workshop: Are you familiar with corpus linguistics?		
• Yes, I am familiar with corpus linguistics (> university studies)	27	10.9%
• No, I am not familiar with corpus linguistics but I have already heard of it (> colleagues, books/articles, conferences, etc.)	24	9.7%
• No, I don't know anything about corpus linguistics	197	79.4%
2) After the workshop I: Do you think that teachers and/or learners may profit from corpus data?		
• Yes, both teachers and learners	32	12.9%
• Yes, but only teachers	208	83.9%
• No	8	3.2%
3) After the workshop II: In which particular fields would you consider consulting or using corpus data in the future? (<i>multiple answers possible</i>)		
• Creation of concordance-based teaching material (> teaching of collocations, patterns, spoken/written differences, etc.)	212	85.5%
• Correction of classtests (> acceptability/idiomaticity of collocations, patterns, phrases, etc.)	137	55.2%
• Creation of word/phrase lists for individual text collections (> set books, texts in 'bilingual subjects' such as history and geography in English medium, etc.)	130	52.4%
• (Other teacher-centred activities)	128	51.6%
• Learner-centred activities (> consultation of corpus data, small-scale corpus studies, corpus browsing, large-scale term papers, etc.)	29	11.7%

2 The Role of Corpus Linguistics in the English Classroom in Germany: What Survey Data Show

Some of the questions that the teachers were asked before and after participating in one of the eight test workshops and the answers they gave are listed above in Table 1.

Before the workshop on applied corpus linguistics, the participants were asked about their previous knowledge about corpus linguistics in general. The answers to the first question in Table 1 paint a bleak picture: some 80% of all qualified English teachers had not come across corpus linguistics before. Under the assumption that the survey trend is more or less representative, the answer to the very first question nicely illustrates the low extent to which corpus linguistics has so far had an impact on teaching practice in Germany.

After the workshop, all participants were asked to answer several questions on the role that corpus linguistics may play in English language teaching in general and in their own classrooms in particular, including questions 2) and 3) in Table 1 above. The distribution of answers to the second question shows that virtually all participants (i.e. more than 95%) do think that English language teaching may profit in one way or another from the advent of corpora. Note, however, that most teachers would only consider making use of corpus data and corpus-based methods themselves. That learners should have access to corpus data as well is not viewed as a fruitful idea by the vast majority. It should be noted that this bias towards teacher-centred corpus activities holds true for the majority of participants in all the test workshops, regardless of whether the focus was more on teacher-centred or on learner-centred techniques. In a sense, this sheds light on an important clash between applied corpus-linguistic research and the average teacher's point of view; while in applied corpus linguistics, there is an increasing tendency to focus on corpus-based activities carried out by increasingly autonomous learners (see Bernardini 2000; Gavioli 2001), most teachers think that corpus data are particularly useful for themselves. This is corroborated by the answers to the third question in Table 1. In the test workshops, a wide range of language-pedagogical applications of corpora were introduced and exemplified – from teacher-centred activities such as the creation of concordance-based teaching material, as described by Flowerdew (2001), Granger and Tribble (1998) and many others, to learner-centred activities such as serendipitous corpus browsing, as sketched out by Bernardini (2000) and others. At the end of the workshops, the participants were supposed to list those activities that they found particularly useful and that they would intend to put into practice in their own classrooms. The important point here is that most teachers, in answering the third question, exclusively focused on teacher-centred activities and showed that learner-centred activities would presumably have no place in their classrooms.²

In conclusion, the results of the survey show quite clearly that the use of corpora, which may have become mainstream in English linguistics, is so far not at all central to the practice of English language teaching in Germany. On the

contrary, only a tiny fraction of English language teachers actually know of the existence of corpus linguistics in the first place.

Paradoxically, most of the teachers who took part in the survey admitted using corpus-based resources, especially corpus-based dictionaries. When they were asked which monolingual English dictionary they tend to use for reference purposes, some 80% listed one of the following corpus-based dictionaries: *Collins COBUILD English Dictionary*, *Oxford Advanced Learner's Dictionary*, *Longman Dictionary of Contemporary English* and *Cambridge International Dictionary of English*. This finding indicates that, on the one hand, most teachers do use corpus-based dictionaries but that, on the other hand, they are not aware of the corpus-linguistic background of these products, i.e. the fact that these dictionaries are based on the quantitative analysis of large and representative samples of naturally occurring language. The same holds true, by the way, for corpus-based learner grammars such as Ungerer's (1999) *Englische Grammatik Heute*. This grammar is increasingly used by both teachers and learners alike in Germany, but the author's comments in the preface on the role of the British National Corpus as a major database of this grammar usually go unnoticed. Here we thus encounter a second gap between corpus-linguistic research and teaching practice. Without any doubt, corpus-based insights into actual language use have already exerted an enormous influence on dictionaries, grammars and modern textbooks that are used by teachers and learners. However, most teachers do not know that many differences between these modern materials and older ones are caused by corpus data and their implications for language teaching. For example, most teachers were surprised when they were told in the workshop that the order in which irregular verbs are taught in modern teaching materials in Germany is largely based on corpus findings, especially those presented by Grabowski and Mindt (1995).

The gap between the rapid development of applied corpus linguistics and its influence on modern classroom resources on the one hand and the average English teacher's knowledge on the other can only be bridged if many more English language teachers are systematically familiarized with the basic foundations, implications and applications of corpus linguistics. This brings me to the need for a large-scale popularization of corpus linguistics among English teachers in Germany (and probably elsewhere too). Most importantly, it is obvious that learners will only get access to corpus data if teachers themselves work with corpora and make them available to their students.

3 The Need for Popularization

According to Aston (2000), there are three fields in which corpus data prove relevant to English language teaching:

- teaching about corpora, as corpus linguistics finds its way into university linguistics curricula;
- exploiting corpora to teach languages, linguistics, and potentially other subjects;
- teaching to exploit corpora, so that learners can explore them for their own purposes. (Aston, 2000: 7)

However, whether or not corpus linguistics is really about to ‘find its way into university curricula’ is open to discussion – at least when it comes to Germany. Even today, it is still perfectly possible for each and every student of English language and literature in virtually all English departments in Germany to take a university degree without ever having delved into corpus linguistics. Thus, it is important to keep in mind that for the time being – and in the foreseeable future – most newly-fledged English teachers enter schools with anything but a detailed knowledge about corpus linguistics. What is more, if most teachers lack this knowledge, they cannot be expected to *exploit corpora to teach languages* nor to *teach [their students] to exploit corpora*.

In the light of the fact that university curricula usually do not include an obligatory corpus-linguistic module, a promising short-term solution to this problem is to offer introductory workshops for qualified English language teachers. The test workshops in which the participants in the survey on which I reported in sections 1 and 2 took part are examples of such a ‘quick-and-dirty’ introduction to corpus linguistics for qualified English language teachers. If, however, the target audience of such workshops are qualified and experienced English language teachers – and not, say, students of English language and literature – it is of paramount importance to offer teachers realistic and easily applicable corpus-based solutions to significant problems that they have been facing in their classrooms. I would contend that the use of corpus data only becomes popular if teachers immediately see the advantage of using corpus data in order to solve existing problems. Involving learners in corpus-based activities continues to be a vital objective, but I would regard it as a second step which teachers will only take after being convinced of the usefulness of corpus data for solving their own teaching problems. In picking up on Aston’s (2000) systematization above, I have already outlined elsewhere (see Mukherjee 2002: 118) that it is the teachers to whom particular attention should be paid in this process of popularization.

As shown in Figure 1, teachers have to be trained in applied corpus linguistics first because only they can be expected to introduce corpus-linguistic methods in the classroom and to involve learners in corpus-based activities. The ultimate objective remains, of course, to make learners work with corpora autonomously.

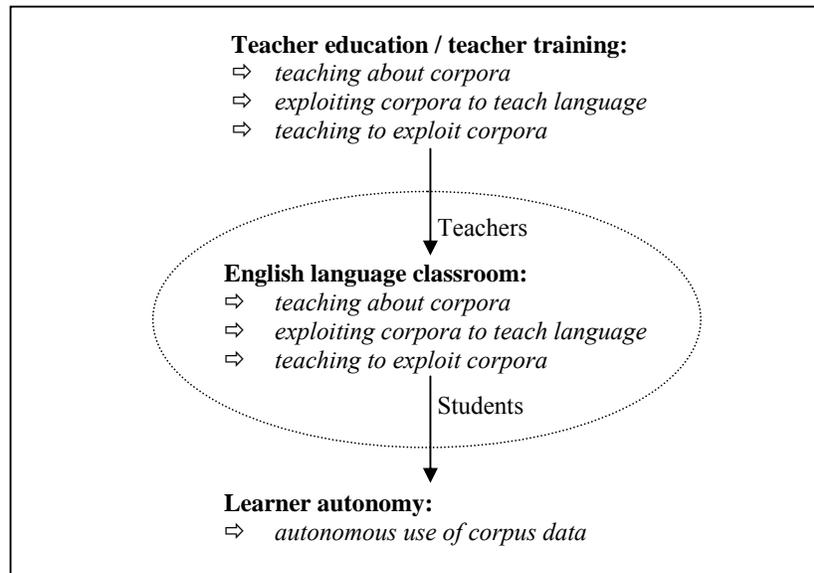


Figure 1: From corpus-experienced teachers to autonomous learners

There is no point in ignoring the fact that most teachers have no prior knowledge about corpus linguistics. Any effort to popularize the language-pedagogical use of corpus data can thus be successful only if we re-focus on such teachers' preconceptions and needs. To this end, I would now like to briefly turn to some aspects of a workshop program for qualified English teachers in more detail – aspects that most participants in the eight test workshops found particularly useful and motivated them to get involved with corpus-linguistic methods. In due course, it is intended to officially include this workshop on corpus linguistics in the teacher training programme which is offered by one of the local state teaching boards in North Rhine-Westphalia. In principle, this institutionalized workshop will then be open to any qualified English language teacher.⁴

4 Refocusing on Teachers' Preconceptions and Needs

The one-day workshop on corpus linguistics will consist of three parts which mirror Aston's (2000) systematization. As shown in Table 2, the focus of the first part – 'teaching about corpora' – is not only on some basic issues of corpus linguistics but also on corpus-based findings that even experienced English language teachers find surprising. This part is thus called the *surprise-the-teacher* module. The eight test workshops (see sections 1 and 2) have shown that this approach makes teachers want to learn more about corpus linguistics right from

the beginning. For example, some 90% of all English language teachers mark the following sentence as wrong because it violates, in their view, the school-grammar rule which states that there should be no *would* in *if*-clauses:

- (1) "I would be grateful if you would send me more specific information."
 => marked as wrong: 221 (89.1%)
 => not marked: 27 (10.9%)

There are many other examples of corpus-based findings that call into question the way English language teachers go about correctness in learner language. Specifically, the discussion of such examples makes it clear to all teachers that their own intuition is often at odds with linguistic reality.

As shown in Table 2, the second module is about *exploiting corpora to teach language*. The approach here is called *help the teacher* because special emphasis is placed on practical problems that virtually all teachers have to face. Picking up on the issue of correctness and correction, corpus data are shown to be useful resources for the teacher because they, for example, provide information on whether particular phrases are idiomatic and instantiate native speakers' "preferred ways of putting things" (see Kennedy 1992: 335) or not.

Table 2: Modules of a one-day workshop on corpus linguistics for qualified English language teachers: an overview

	Aspect	Approach
Module I	<ul style="list-style-type: none"> • <i>teaching about corpora</i> ⇒ basic notions: corpus design, major corpora, authenticity, representativeness etc.	<ul style="list-style-type: none"> • "surprise the teacher" ⇒ corpus-based findings that run counter to preconceived ideas: e.g., <i>would</i> in <i>if</i> -clauses
Module II	<ul style="list-style-type: none"> • <i>exploiting corpora to teach language</i> ⇒ idiomaticity, native-like selection, spoken vs. written English, genre differences etc.	<ul style="list-style-type: none"> • "help the teacher" ⇒ using corpus data to solve teaching problems: e.g., correction of class tests
Module III	<ul style="list-style-type: none"> • <i>teaching to exploit corpora</i> ⇒ learner autonomy, data-driven learning, media literacy etc.	<ul style="list-style-type: none"> • "pass it on to the learner" ⇒ involving learners: e.g., identification of genre-specific realisations of moves

Many examples of such usage problems, especially in written school work, are discussed in this part of the workshop. Another aspect that is covered in this section is the corpus-based teaching of spoken English. In Germany, many colleagues use the derogatory term *Abiturspeak* – with *Abitur* being the German A-levels – to refer to the phenomenon that many advanced learners leave school without being sufficiently able to use natural spoken English:

Leider ist das in den Klassenzimmern anzutreffende Englisch in der Regel die geschriebene Sprache, mündlich angewendet. [Unfortunately, it is written English, used in the spoken medium, that we usually encounter in the classroom.] (Kieweg 2000: 8; my translation)

In fact, learners very often speak just as they write. Many teachers are aware of this problem, and in an institutionalized workshop on corpus linguistics, it should be our intention to capitalize on their classroom experience and show them how they can use corpus data in order to identify, for example, frequently occurring spoken items and patterns. The principal objective of this module is, of course, to provide teachers with hands-on practical experience so that they regard corpus data not just as a recent (but useless) trend in language-pedagogy but as a helpful, problem-solving resource.

It is only in the last module (see Table 2) that the emphasis will be shifted to learners' interaction with corpora. As pointed out in section 2, most teachers remain sceptical about learner autonomy in this field, and the only thing that we aim at in this last section is to provide some sort of topic-opener in this regard. However, it should be noted that even among the sceptical majority of teachers some applications turn out to be more convincing than others. For example, Henry and Roseberry's (2001) corpus-based genre approach to language teaching is a method that some participants in the test workshops have already tried out in their own classrooms; this method is therefore a good candidate for inclusion in the third module of an institutionalised workshop.⁵

5 Concluding Remarks

I hope to have shown that many qualified English language teachers in Germany do not know very much, if anything at all, about the rapid developments in corpus linguistics and its language-pedagogical applications. Against this background, I have tried to sketch out how the use of corpus data may become more popular among teachers in the German context. Let me emphasise once again that there is, at present, a large gap between the wealth of applied corpus-linguistic research and the teaching practice in Germany which so far has only been affected to a very limited extent by this research. Closing this gap is a challenge to applied corpus linguists and, perhaps more importantly, to those who are involved in teacher training (both for trainee and qualified teachers). In trying to meet this

challenge, special emphasis should be placed on the average teacher's pre-conceptions and practical needs. For, as Kettemann (1997: 70) correctly points out, it is only by updating teachers' *brainware* that we can change teaching practice in the English language classroom.

I should think that the kind of institutionalized workshop that is envisaged in the present paper would help to popularize corpus-based methods in the English classroom not only in Germany but also in other countries with English as a foreign language (EFL). While the overall modular design may be picked up on in virtually all EFL countries, some aspects would need to be adapted to each individual country. For example, it would be useful to take into account the typical learner errors that are caused by structural differences between the learners' native language and English and to focus on corpus-based methods that may help to iron out those typical cross-linguistic interferences. Also, it is quite clear that the kind of workshop suggested in the present paper is based on the language-pedagogical concepts of authentic language use, inductive learning and learner autonomy. While the corpus-based, data-driven approach to language learning is perfectly in line with English curricula in Germany, one would need to modify the workshop if curricular frameworks for English language teaching in other EFL countries are fundamentally different (e.g., by emphasizing written language use and deductive language learning).

Notwithstanding these caveats, corpus linguistics will find its way into the reality of English language teaching in all EFL countries only if not only students of English language and literature but also qualified English teachers are trained on the job. The institutionalisation of introductory workshops may offer a way forward from the present gap between applied corpus-linguistic research and the reality of English language teaching.

Notes

1. Whether the population of 248 teachers is a truly representative sample of the entirety of all English teachers is, of course, open to debate. However, since the teachers were randomly selected, the general trends are, in my view, indicative of similar trends in the whole teacher population. There is no doubt that further research, including longitudinal studies, is needed.
2. It should be noted in passing – and this does not come as a surprise – that there is a significant correlation between the age group of the participants and their willingness to let their students work with corpora autonomously. But since the average age of secondary school teachers in Germany is just below fifty, it goes without saying that most teachers belong to the group that is rather sceptical about learner-centred activities. For example, only 3 of 98 teachers of 50 to 65 years of age (3.1%) mentioned learner-centred activities in answering the third question in Table 1, while 25 of 46 teachers of up to 30 years of age

(54.3%) did. Unsurprisingly, too, 23 of the 27 teachers (85.2%) that had already been familiar with corpus linguistics before taking part in the test workshop were 30 years of age or younger. No-one in the 50+ age-group, on the other hand, considered himself/herself to be already *familiar with corpus linguistics*.

3. In fact, most of my students in Giessen and – until recently – in Bonn are not very keen on linguistic branches that make use of computers; I agree with Seidlhofer (2000: 208) that “most of our undergraduates are genuinely technophobic.” This negative attitude towards the computer-based description and analysis of language does not usually change once these students have obtained their degree and become trainee teachers and – eventually – qualified teachers.
4. In this context, I am particularly grateful to Jan-Marc Rohrbach for sharing – and discussing – with me his classroom experience and to Kunibert Broich for helping to pave the way for an institutionalisation of such a workshop on corpus linguistics.
5. In most cases, however, the teacher remains strongly involved in the corpus-based activities and we can thus not speak of true learner autonomy, as for example Rohrbach’s (2003) illuminating report on a corpus-based genre approach to the production of travel brochures in class 9 shows. Nevertheless, the workshop is considered to be more than successful if teachers are enabled – and willing – to work with corpora themselves, which is a prerequisite for corpus-based activities on the part of the learners somewhere down the line.

References

- Aston, G. (2000), Corpora and language teaching, in L. Burnard and T. McEnery (eds), *Rethinking language pedagogy from a corpus perspective: Papers from the Third International Conference on Teaching and Language Corpora*, Frankfurt am Main: Peter Lang, pp. 7-17.
- Aston, G. (ed.) (2001), *Learning with corpora*, Houston, TX: Athelstan.
- Bernardini, S. (2000), Systematising serendipity: Proposals for concordancing large corpora with language learners, in L. Burnard and T. McEnery (eds), *Rethinking language pedagogy from a corpus perspective: Papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, pp. 225-234.
- Burnard, L. and T. McEnery (eds) (2000), *Rethinking language pedagogy from a corpus perspective: Papers from the Third International Conference on Teaching and Language Corpora*, Frankfurt am Main: Peter Lang.
- Flowerdew, L. (2001), The exploitation of small learner corpora in EAP materials design, in M. Ghadessy, A. Henry, and R.L. Roseberry (eds), *Small*

- corpus studies and ELT: Theory and practice*, Amsterdam: John Benjamins, pp. 363-379.
- Gavioli, L. (2001), The learner as researcher: Introducing corpus concordancing in the classroom, in G. Aston (ed.), *Learning with corpora*, Houston, TX: Athelstan, pp. 108-137.
- Grabowski, E. and D. Mindt (1995), A corpus-based learning list of irregular verbs in English, *ICAME Journal* 19: 5-22.
- Granger, S. (ed.) (1998), *Learner English on computer*, London: Longman.
- Granger, S. and C. Tribble (1998), Learner corpus data in the foreign language classroom: Form focused instruction and data-driven learning, in S. Granger (ed.), *Learner English on computer*, London: Longman, pp. 199-209.
- Henry, A. and R.L. Roseberry (2001), Using a small corpus to obtain data for teaching a genre, in M. Ghadessy, A. Henry, and R.L. Roseberry (eds), *Small corpus studies and ELT: Theory and practice*, Amsterdam: John Benjamins, pp. 93-133.
- Johns, T. (1991), Should you be persuaded: Two examples of data-driven learning materials, *English Language Research Journal*, 4: 1-16.
- Kennedy, G. (1992), Preferred ways of putting things with implications for language teaching, in J. Svartvik (ed.), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82*, Berlin: Mouton de Gruyter, pp. 335-373.
- Kettemann, B. (1997), Der computer im Sprachunterricht, in M. Stegu and R. de Cilia (eds), *Fremdsprachendidaktik und Übersetzungswissenschaft: Beiträge zum 1. verbal-workshop, Dezember 1994*, Frankfurt am Main: Peter Lang, pp. 63-72.
- Kieweg, W. (2000), Zur Mündlichkeit im Englischunterricht, *Der fremdsprachliche Unterricht Englisch* 34 (5): 4-9.
- Mukherjee, J. (2002), *Korpuslinguistik und Englischunterricht: Eine Einführung*. Frankfurt am Main: Peter Lang.
- Rohrbach, J-M. (in press), Don't miss out on Göttingen's nightlife: Genreproduktion im Englischunterricht, *Praxis des neusprachlichen Unterrichts*, 50.
- Seidlhofer, B. (2000), Operationalizing intertextuality: Using learner corpora for learning, in L. Burnard and T. McEnery (eds), *Rethinking language pedagogy from a corpus perspective: Papers from the Third International Conference on Teaching and Language Corpora*, Frankfurt am Main: Peter Lang, pp. 207-223.
- Sinclair, J.M. (ed.) (1987), *Looking up: An account of the COBUILD project in lexical computing*, London: Collins.
- Tribble, C. (2000), Practical uses for language corpora in ELT, in P. Brett and G. Motteram (eds), *A Special interest in computers: Learning and teaching with information and communications technologies*, Whitstable, Kent, UK: IATEFL, pp. 31-41.
- Ungerer, F. (1999), *Englische grammatik heute*, Stuttgart, Germany: Ernst Klett.