

UTE RÖMER (HANOVER)

## **Learner Language and the Norms in Native Corpora and EFL Teaching Materials: A Case Study of English Conditionals**

### **1. Introduction**

In the determination of standards and norms, language corpora (i.e. large computerised collections of spoken and/or written text used for linguistic analysis) can be considered invaluable tools. They enable the researcher to describe in detail how certain linguistic phenomena are typically used by different groups of speakers or writers in different contexts. They also help us determine whether the usage norms are in line with the rules we find in prescriptive accounts of the language, or whether there are any conflicts between what Mair and Mollin (this volume) label "standard-as-construct" and "standard-in-use".

The present paper will take a corpus approach to lexicalgrammatical norms in three types of language data: (i) English as a Foreign Language (EFL) textbook English, (ii) British native speaker English, and (iii) advanced German learner English. By means of a comparative case study of conditional *if* (as in *This might happen if you give money away to members of your family*; BNC\_written) in these different kinds of English, the paper will investigate whether (and if so how) the norms of actual native speaker usage differ from the norms put forward in EFL teaching materials. The paper will also examine which norms German learners tend to follow. The evaluation of the corpus data will then lead to a discussion of aspects that need to be considered in the determination of the most appropriate norm (or norms) for English language learning and teaching.

### **2. Language Norms in Different Sets of Data: The Case of If-clauses**

Do different types of language data display different norms – and, if so, in what ways do they differ? Figure 1 provides an overview of the types of data and corpora that have been selected in order to address this question. For the analysis of EFL textbook English, a corpus that consists of spoken-type texts was used (i.e. texts that are supposed to represent speech, e.g. dialogues and interviews). This corpus, the German English as a Foreign Language Textbook Corpus (GEFL TC; cf. Römer 2004b; 2005), contains material from twelve volumes of two of the most widely-used textbook series in German grammar schools: *Green Line New* (Klett) and *English G 2000 A* (Cornelsen). GEFL TC was originally compiled as part of a study of progressives in real spoken English and in 'school' English (cf. Römer 2005). Spoken-type texts were selected to ensure comparability with data collected from the spoken components of the British National Corpus and the Bank of English. The textbook corpus represents a

considerable part of the input that language learners actually get in the first six years of English instruction (termed "real learner input" in Römer 2004b).

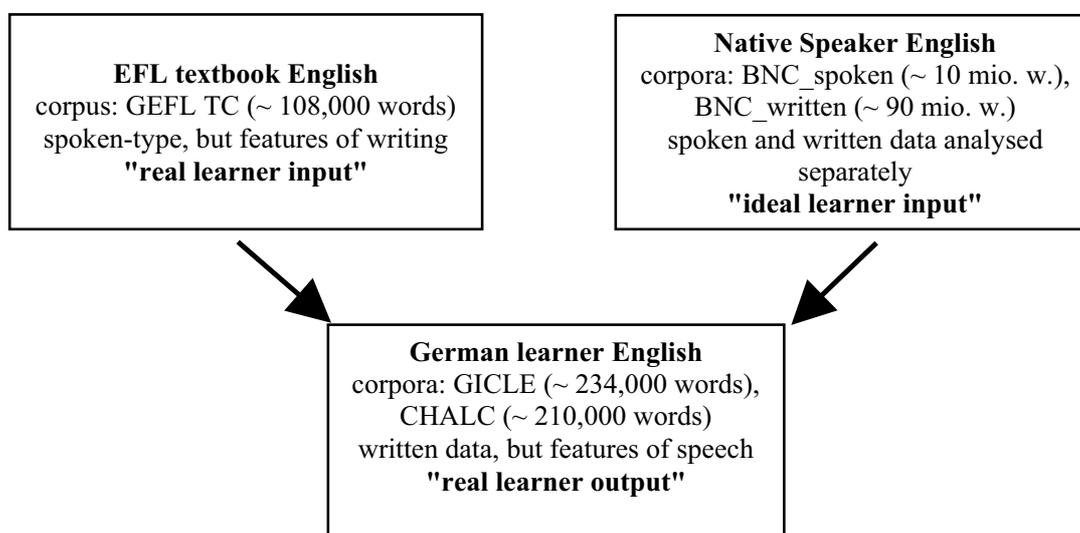


Figure 1: Selected types of data and corpora

Even though the texts included in GEFL TC aim at representing spoken English, several of the findings discussed in Römer (2005) indicate that this spoken-type coursebook English in many respects resembles writing more than speech. This confirms Lewis' (1993, 8) observation that "[m]any of the dialogues in textbooks are much closer stylistically to written, rather than spoken English". I hence assumed that the textbook norm would lie somewhere between speech and writing, and decided to include spoken and written data in the description of native speaker English (referred to as "ideal learner input" in Römer 2004b).<sup>1</sup> Two random sets of 100 instances of *if* were collected from the 10 million word spoken and from the 90 million word written component of the British National Corpus (henceforth BNC\_spoken and BNC\_written). From these four concordance sets, those lines were deleted in which *if* was not used in its conditional sense (e.g. *I just like to know if Gordon is actually aware of the terms*; BNC\_spoken). The discussion below is based on the filtered datasets.

The same applies to the four sets of *if*-concordance lines (also 100 examples each) retrieved from the two selected learner corpora: CHALC and GICLE. Both corpora contain written language output of advanced German learners of English. While CHALC, the Cologne-Hanover Advanced Learner Corpus, consists of (mainly linguistic) term papers and essays by 2<sup>nd</sup> to 5<sup>th</sup> year students of English from Hanover University and Cologne University (currently making up roughly 210,000 words),<sup>2</sup> GICLE, the German part of the International Corpus of Learner English (Granger et al. 2002), comprises mainly non-academic argumentative essays written by 3<sup>rd</sup> and 4<sup>th</sup> year students from the universities of Augsburg, Basel, Dresden, and Salzburg (cf.

1 The native speaker norm is also described as an ideal input norm by Mukherjee (2005, 15).

2 CHALC is designed as a monitor corpus, which means that texts are being added on a regular basis.

Nesselhauf 2005, 44) and has a size of about 234,000 words.<sup>3</sup> We are here dealing with a type of "real learner output" that, like the data from GEFL TC, shows features of both speech and writing. To quote Granger (forthc.), advanced learner writing, as covered by GICLE and CHALC, is "often characterised by an overly spoken style". It will now be interesting to see what kinds of if-clauses are mainly used in the different corpora and how the datasets relate to each other in terms of grammatical norms. We will start with a discussion of if-clauses in EFL textbook English and focus on the types of conditionals that are put forward as a model in pedagogical descriptions.

### 2.1. EFL Textbook English: The Teaching Norm

EFL grammar books usually distinguish between three types of if-clauses. Conditional "type 1" serves to express "real or open conditions", whereas conditionals "type 2" and "type 3" refer to "unreal conditions" (Fleischhack et al. 2001, 186). Examples of the three types, all taken from GEFL TC, are given in (1) to (3) below, together with the tense form sequence used in each type.

- (1) *If you eat your hat, you'll be ill.* (= type 1; if-part: simple present – main part: *will* + infinitive)
- (2) *But if you had a job on land, you would see your family more often.* (= type 2; if-part: simple past – main part: *would* + infinitive)
- (3) *If they had been allowed to choose, they would have spent this vacation on a beach.* (= type 3; if-part: past perfect – main part: *would* + *have* + past participle)

In order to determine the teaching norms that apply to if-clauses, all conditionals in the textbook corpus (211 examples altogether) were classified according to their combination of tense forms in if-part and main part. Figure 2 displays the results of the tense form assignment. We see that most of the 211 if-clauses in GEFL TC (72.7%) follow the EFL grammar rules and can be assigned to one of the three types mentioned above (type 1: 39%; type 2: 25.1%; type 3: 8.6%). Most frequent is the "real condition" type, followed by the first of the "unreal condition" types (type 2). With 18.7% the third most common conditional in GEFL TC is a type that tends to be marginalised in school grammars and treated as a subtype of type 1 (cf. Fleischhack / Schwarz 2001, 188). This type has a simple present form in both parts (e.g. *If you help us, this is yours.* GEFL TC) and is used to state general validities. Further tense form sequences that occur in GEFL TC, such as mixes of types 2 and 3 (e.g. *If I had big ones [muscles] like the Malleys, I'd never have been able to get through that hole in the fence*), were summarised under "other combinations". These other combinations are all very rare in the textbooks. It can thus be said that the focus in teaching materials is on if-clause types 1-3 which serve as models for learners.

---

3 For more information on the ICLC project, the reader is referred to Granger (ed.) (1998) and Granger (2002).

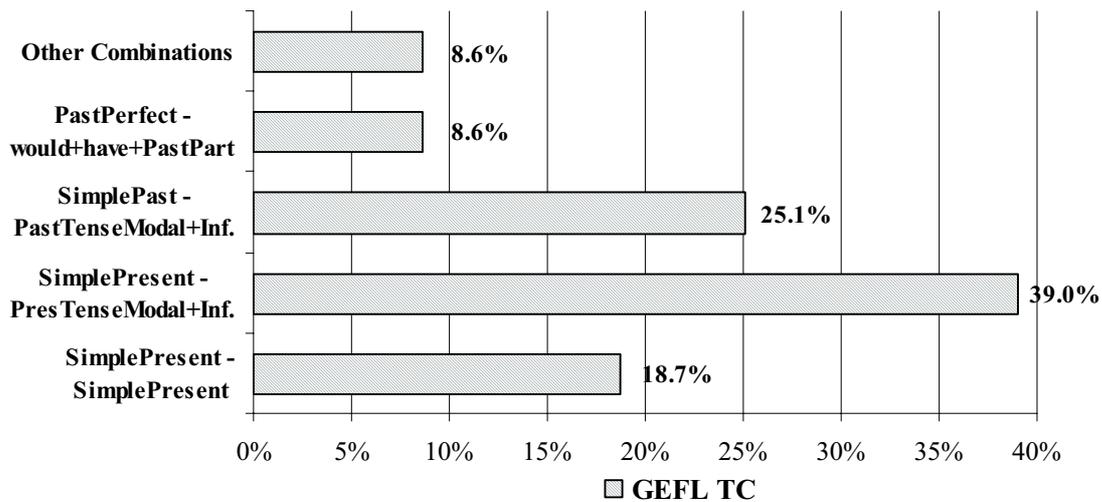


Figure 2: Tense form sequences in if-clauses in GEFL TC

## 2.2. British Native Speaker English: The Usage Norm

Let us now turn to an analysis of if-clauses and their most common tense form sequences in spoken and written native speaker English. The graph in Figure 3 serves to illustrate how the usage norm relates to the EFL teaching norm discussed in section 2.1 of this paper.

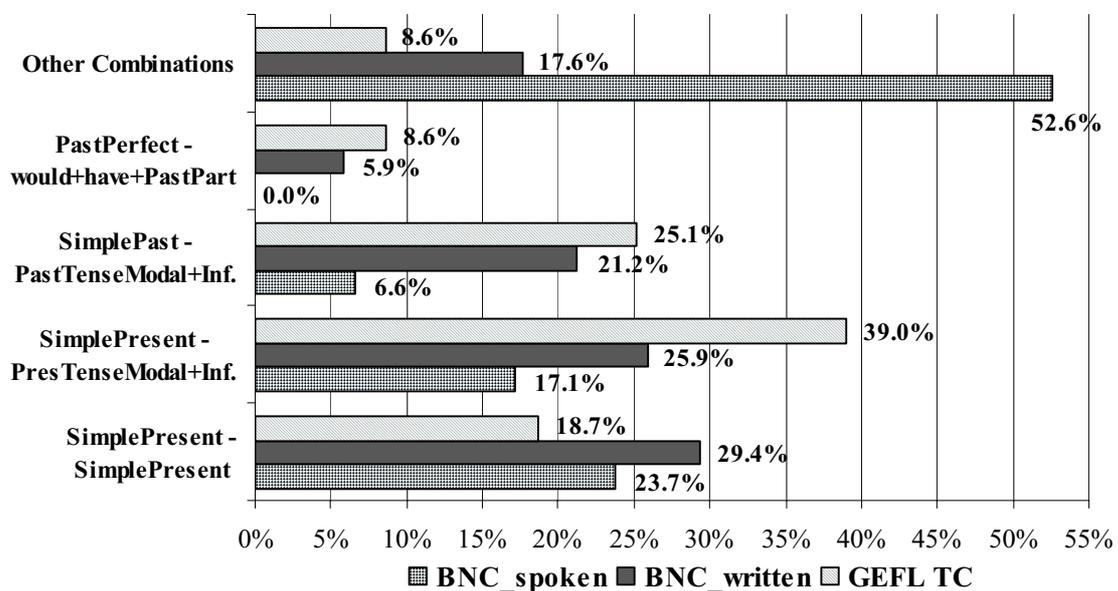


Figure 3: Tense form sequences in if-clauses in BNC\_spoken and BNC\_written (compared to GEFL TC)

We observe not only differences between speech and writing, but also between real English and textbook English. If-clause types 1-3 are less frequent in actual language use than the teaching norm suggests. Conditionals of the first type (simple present – present tense modal + infinitive), for example, make up 39% of the GEFL TC dataset and only 17.1% of the analysed BNC\_spoken and 25.9% of the BNC\_written data. On

the other hand, many of the if-clause types that were found in the real English data (summarised under "other combinations") do not occur in the textbooks at all, and the most common type (simple present in both parts of the if-clause) in the BNC\_written and BNC\_spoken (29.4% and 23.7% respectively) is comparatively rare in GEFL TC. Figure 3 also shows that, although GEFL TC only contains representations of speech, the textbook distribution is closer to the written than the spoken usage norm.

On the whole we can say that the usage norm allows for a variety of tense form sequences in if-clauses (many of which are not included in the analysed EFL teaching materials), with combinations of simple present + simple present being the most frequent type, followed by type 1 conditionals, such as [...] *demand will materialise if a suitable product and marketing mix are introduced* (BNC\_written).

### 2.3. German Learner English: Does It Mirror Any Norms?

Now that we know how if-clause types are distributed in real English and 'school' English, it will be interesting to see how advanced German learners use conditionals, and to examine whether learner English rather depends on the teaching or on the usage norm. Since many of the learners who contributed to CHALC and GICLE spent some time abroad in an English speaking country between leaving school and entering university, a certain influence of the native speaker norm can be expected to surface in the data.

In Figure 4 the shares of tense-form combinations in conditionals in GICLE and CHALC are displayed and compared with the GEFL TC and BNC results discussed above. The picture we get there is rather complex and somewhat confusing. We notice differences between the two learner corpora, between the learner corpora and the textbook corpus, and between the learner and native speaker corpora. By just looking at the relative frequencies given in Figure 4 we cannot clearly determine whether, in terms of if-clause use, advanced German learner English is closer to the teaching or the usage norm. A closer examination of the GICLE and CHALC datasets, however, provides us with some interesting insights into the problems learners have, and offers possible explanations for the over- and under-representation of some tense form sequences in the learner corpora conditionals.<sup>4</sup>

The very high shares of if-clauses (62.5% and 50%) with a simple present form in both parts, for instance, are probably due to the types of texts included in GICLE (argumentative essays) and CHALC (mainly linguistic term papers). In the GICLE essays and CHALC term papers, we find a large number of examples of the type given in (4) to (7), which is perhaps not very surprising.

(4) [...] *the final syllable is stressed if it is heavy.* (CHALC)

---

4 To get a less distorted picture, it would have been useful to include data from a spoken learner corpus, e.g. the German component of LINDSEI (the Louvain International Database of Spoken English Interlanguage (<http://cecl.fltr.ucl.ac.be/Cecl-Projects/Lindsei/lindsei.htm> (1/12/2006)); see also Brand et al. 2006). Unfortunately, at the time of data collection for this study, I did not have access to such a corpus.

- (5) *B m-commands C if B does not dominate C.* (CHALC)
- (6) *If you drive 120 km per hour, everybody overtakes you.* (GICLE)
- (7) *If it comes down to cars, German men are not longer able to use their common sense any more.* (GICLE)

Also worth examining was the unexpectedly large group (30.06%) of "other combinations" conditionals from GICLE. What we find in this group is not the combinations that occur in the BNC\_spoken dataset, but many odd if-clauses with tense-form sequences that are not (or hardly ever) used by native speakers (see examples (8) and (9) below). A number of odd if-clauses (judgements confirmed by two native speakers) could also be found in the "type 1" and "type 2" groups. Two illustrative examples from CHALC are displayed in (10) and (11). In these and other cases it seems that the learners are determined to stick to the rules formulated in EFL textbooks and grammars and produce if-clauses which contain the "correct" sequence of tenses, no matter whether the tense forms suit what they actually want to express.

- (8) *Everything would be changed if one turns Augsburg's city centre into a pedestrianized zone.* (GICLE)
- (9) *[...] but if this person will join in those new games, he will surely get a lot of fun out of it.* (GICLE)
- (10) *If he was an itinerant preacher, this very fact would be the most significant reason for him to have written a book like Piers Plowman.* (CHALC)
- (11) *If we add a tag question to the sentence, we will get "This essay is obligatory for course participation, isn't it?"* (CHALC)

By way of summing up our observations on if-clauses in GICLE and CHALC, we can say that neither of the norms described in sections 2.1 and 2.2 is clearly mirrored in the type of learner English covered by our corpora, but that the learners, even though they are quite advanced, seem to be confused about the choice of appropriate tense-form sequences in conditionals. A possible explanation for this learner confusion could be that learners find themselves between conflicting values and influences: EFL textbook/grammar influence, real English influence, and the influence of their first language (here German), which may contain potentially concurring structures. The next section will discuss what could be done to improve the situation for the learner, and consider the above findings in the light of an ideal target model for English language teaching.

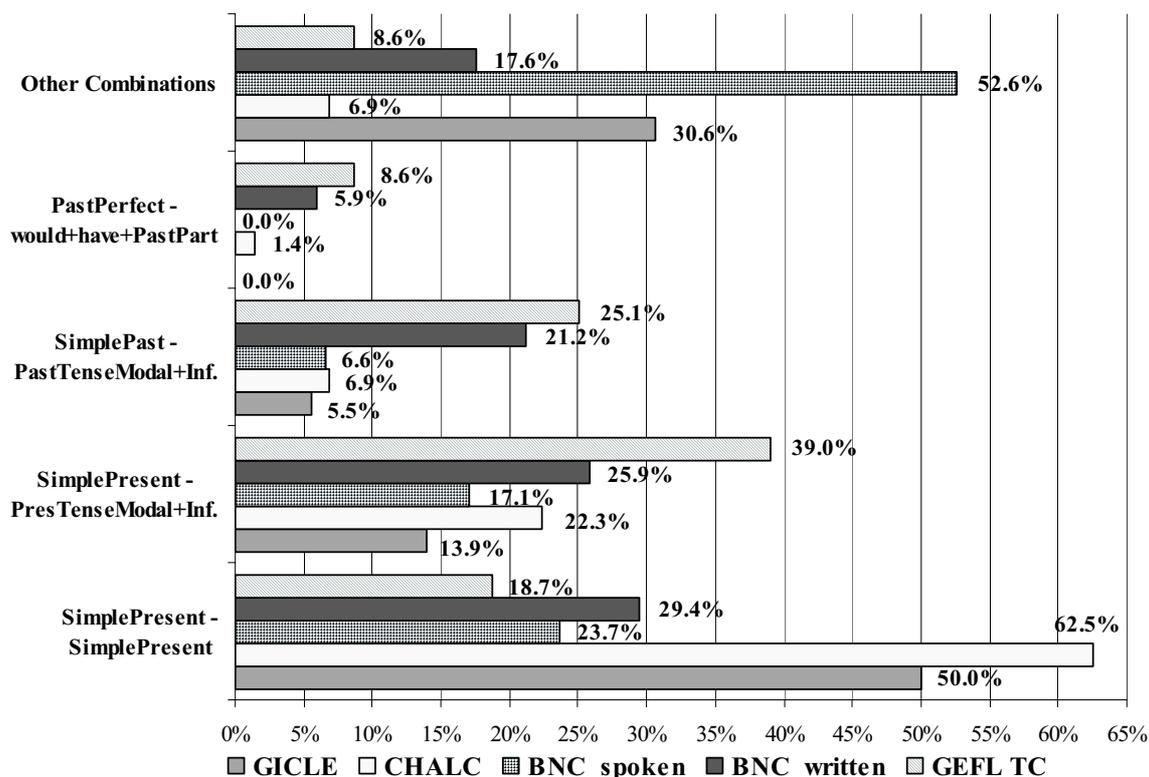


Figure 4: Tense form sequences in if-clauses in GICLE and CHALC (compared to GEFL TC and BNC\_spoken/written)

### 3. Determining the Most Appropriate Norm for Language Teaching

Like several other corpus studies that compare selected language phenomena in native speaker English with the presentation of the same points in EFL coursebooks and grammars (cf. e.g. Conrad 2004; Grabowski et al. 1995; Mindt 1997; Römer 2004a; Schlüter 2002), our analysis of conditionals in different corpora has highlighted mismatches between English in use and English 'in the books'. If we agree that learners ought to be prepared for a participation in real communicative situations and presented with a genuine language model in the classroom, such findings may cast doubt on the status of the norms put forward in EFL teaching materials.

We could hence argue that, since the current teaching norm clashes with the usage norm in several respects, it is perhaps not the most appropriate one for our learners, and suggest a real English model as an ideal norm instead. The question we then have to ask is what type of real English our descriptions for the learner should be based on. Do we want to stick to a native speaker target norm and aim at accurate learner language, or is our main purpose to help learners become successful communicators in different communicative contexts? Perhaps Lewis (1993, vi) is right in saying that "[s]uccessful language is a wider concept than accurate language" and that "[l]anguage is most adequately assessed on a scale of communicatively successful/unsuccessful, rather than right/wrong" (Lewis 1993, 45; cf. also Cook 1999, 185). I assume that for

most of the learners at German grammar schools, a knowledge of appropriate language use and intelligibility by future communication partners would be more sensible and more realistic aims than native speaker accuracy. As Hunston and Francis (2000, 271) note in the context of teaching lexical grammar, "a learner does not need to get all patterns right but needs to get some patterns right, in order to be understood". What our learners need in order to get some patterns right and to use the language appropriately, I would argue, is not necessarily native speaker input but the input of expert language users, and that the equation "native speaker English = ideal learner input" may have to be reassessed. Tribble (1997) clearly has an important point to make when he states that

[. . .] the most useful corpus for learners of English is the one which offers a collection of *expert performances* (Bazerman 1994, 131) in genres which have relevance to the needs and interests of the learners. (Tribble 1997, 3; emphasis in original)

Included in this quote is another crucial aspect that has to be considered when we determine the ideal learner input norm: learners' needs. In the selection of expert models, it makes sense to take into account what the learner needs and which model s/he actually wants to approximate to. With many groups of learners these needs may be hard to determine (see also Gnutzmann, this volume), but in some cases (e.g. with students at a grammar school of economics) it may be possible to predict which discourse communities the learners will be part of and what type(s) of English they will need for successful future communication. An appropriate target norm for the learner could then be derived on the basis of corpora that capture the performance of the discourse communities in question.

#### 4. Conclusion

In this paper the topic of norms in language description and language pedagogy has been approached from a corpus perspective. I have examined the use of if-clauses in corpora of EFL textbook English and native speaker English, and found some deviations between the teaching norm and the usage norm. The subsequent analysis of data from two corpora of German learner English indicated that even advanced learners have problems with the use of if-clauses – perhaps because they are confused by the conflicting input they get. The value of the existing teaching norm was questioned, and it was suggested that teaching materials should be based more on real language, and that the results from corpus work should be taken more seriously.

I hope to have shown that the determination of an ideal norm for language teaching is a complex and complicated task, and that more related studies on lexicalgrammatical phenomena in expert and learner performance data will be necessary to back up my findings. However, if we accept that the most appropriate target norms will depend on which speech communities the learners wish to be part of, and if we take into account what the learners need to know in order to become successful communicators in the respective communities, it should be possible to select appropriate corpora and on their basis develop tailored resources that are better suited to prepare learners for active participation in real-life communication than the materials that are currently being used.

## References

- Brand, Christiane; Kämmerer, Susanne (2006): "The *Louvain International Database of Spoken English Interlanguage* (LINDSEI): Compiling the German Component," in: Braun, Sabine; Kohn, Kurt; Mukherjee, Joybrato (eds.): *Corpus Technology and Language Pedagogy*. Frankfurt a.M.: Peter Lang, 127-140
- Conrad, Susan (2004): "Corpus Linguistics, Language Variation, and Language Teaching," in: Sinclair, John McH. (ed.): *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, 67-85
- Cook, Vivian (1999): "Going Beyond the Native Speaker in Language Teaching," *TESOL Quarterly*, 33.2, 185-209
- Fleischhack, Erich; Schwarz, Hellmut (2001): *Cornelsen English Grammar*. Berlin: Cornelsen
- Grabowski, Eva; Mindt, Dieter (1995): "A Corpus-Based Learning List of Irregular Verbs in English," *ICAME Journal*, 19, 5-22
- Granger, Sylviane (ed.) (1998): *Learner English on Computer*. London: Longman
- Granger, Sylviane (2002): "A Bird's-Eye View of Learner Corpus Research," in: Granger, Sylviane; Hung, Joseph; Petch-Tyson, Stephanie (eds.): *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 3-33
- Granger, Sylviane (2007): "19. Learner Corpora," in: Lüdeling, Anke; Kytö, Merja (eds.): *Corpus Linguistics. An International Handbook*. Berlin: Mouton de Gruyter
- Granger, Sylviane; Dagneaux, Estelle; Meunier, Fanny (eds.) (2002): *International Corpus of Learner English. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain
- Hunston, Susan; Francis, Gill (2000): *Pattern Grammar. A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins
- Lewis, Michael (1993): *The Lexical Approach*. Hove: Language Teaching Publications
- Mindt, Dieter (1997): "Corpora and the Teaching of English in Germany," in: Wichmann, Anne; Fligelstone, Steven; McEnery, Tony; Knowles, Gerry (eds.): *Teaching and Language Corpora*. London: Longman, 40-50
- Mukherjee, Joybrato (2005): "The Native Speaker Is Alive and Kicking: Linguistic and Language-pedagogical Perspectives," *Anglistik*, 16.2, 7-23
- Nesselhauf, Nadja (2005): *Collocations in a Learner Corpus*. Amsterdam: John Benjamins
- Römer, Ute (2004a): "A Corpus-Driven Approach to Modal Auxiliaries and their Didactics," in: Sinclair, John McH. (ed.): *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, 185-199
- Römer, Ute (2004b): "Comparing Real and Ideal Language Learner Input: the Use of an EFL Textbook Corpus in Corpus Linguistics and Language Teaching," in: Aston, Guy; Bernardini, Silvia; Stewart, Dominic (eds.): *Corpora and Language Learners*. Amsterdam: John Benjamins, 151-168
- Römer, Ute (2005): *Progressives, Patterns, Pedagogy. A Corpus-Driven Approach to English Progressive Forms, Functions, Contexts and Didactics*. Amsterdam: John Benjamins
- Schlüter, Norbert (2002): *Present Perfect. Eine korpuslinguistische Analyse des englischen Perfekts mit Vermittlungsvorschlägen für den Sprachunterricht*. Tübingen: Narr
- 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 UP; <http://www.ctribble.co.uk/text/Palc.htm> (22/8/2006)