

Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching

Ute Römer

University of Hanover, Germany

While there is substantial research in the field of corpus linguistics and language teaching based on native-speaker and non-native speaker corpora, the language of EFL teaching materials has not been systematically analysed so far. The present paper discusses the construction and use of an electronic corpus of EFL textbook texts, focussing on the potential of such a corpus in applied corpus linguistics. A case study on if-clauses in spoken English and “school” English demonstrates in which ways analyses based on textbook corpora can lead to valuable insights for linguists and language practitioners and how they may help to improve teaching materials.

1. Introduction

It is widely accepted that native speaker corpora and non-native speaker or learner corpora can be very useful in foreign language learning and teaching. Many linguists value the great impact of these data collections and base their research on the analysis of native speaker and learner *output*. So far comparatively few people have studied learner *input*. I would like to argue that next to analysing the language produced by learners and the language produced by more competent speakers of English it will also be helpful to look at the input pupils actually get in their English lessons. A large part of this language learner input is represented by the textbooks used in teaching English as a Foreign Language (EFL textbooks).¹

In the present paper it will be discussed how an electronic corpus of EFL textbook texts can be used to help us answer two crucial questions related to language teaching: “Do we teach our pupils authentic English, i.e. do we con-

front them with the same type of English they are likely to be confronted with in natural communicative situations?" and "What can we do to improve EFL teaching materials?". The creation and the design of one particular corpus, consisting of texts taken from German EFL textbooks, will be described. To provide evidence of the usefulness of such a corpus for researchers, the results of an empirical study on if-clauses will be presented and evaluated with respect to their significance for English Language Teaching (ELT).

2. The linguistic potential of an EFL textbook corpus

The first question which will be dealt with here is "Why do we need an EFL textbook corpus?". To start with, an EFL textbook corpus can obviously be used to systematically analyse EFL textbook language, i.e. the kind of language pupils are confronted with when learning English. Also, as the title of this paper suggests, such a corpus is what we need if we want to compare what some people term "school English" (cf. e.g. Mindt 1996:232) with authentic or "real" English as represented in a reference corpus; or, in other words, if we want to compare *real* language learner input with what I would like to call *ideal* learner input.

Like any other computerised corpora, an electronic corpus of EFL textbook texts can be used to calculate frequencies of occurrence of single lexical items or word combinations (phrases, multi-word items).² It also allows a closer analysis of contextual phenomena and provides answers to questions like "In which contexts are certain lexemes presented in the textbooks (as opposed to native speaker corpora)?" and "Do we find the same collocational patterns in EFL textbooks that we find in a native speaker corpus?". Different meanings of polysemous words or structures can be found and analysed, and the lexical-grammatical progression in textbooks can be traced. Finally, a textbook corpus may yield more about the status of authenticity in English language teaching – an issue we will briefly deal with in the next section.

3. Authenticity in the classroom – a controversial topic

The problem of authenticity in ELT has been discussed for many years in numerous scholarly books and articles (e.g. Amor 2002, Breen 1985, Taylor 1994) and the debate has recently resurfaced (cf. Widdowson 2000 and Stubbs

2001). What authenticity really means in a language teaching context, which different types of authenticity play a role, and whether or not we want to teach authentic English to our pupils are highly controversial questions among linguists and didacticians. Basically, the discussion centres around the question "Should English produced in natural communicative situations form the basis of our teaching or should we use invented texts and examples specifically created for the purpose of teaching in our course materials?". It may thus be interesting to see what the situation is like at present and how natural the language which learners are presented with in the EFL classroom really is.

After an analysis of several EFL textbooks used in German grammar schools my observation was that what we tend to find in these books is a simplified, non-authentic kind of English. Pupils are mainly presented with invented sentences, sentences which probably have not occurred in any natural speech situation before (and which probably never will). The short dialogue in (1), from an introductory EFL textbook, may serve to illustrate this phenomenon.

- (1) MR SNOW Hello, Wendy.
 MRS SNOW Hello, Ron.
 MR SNOW Where are the girls?
 Are they packing?
 MRS SNOW Yes, they are.
 MR SNOW Or are they playing?
 MRS SNOW No, they aren't, Ron. They are packing.
 (Schwarz 1997:45)

It is rather doubtful whether texts like this can better serve the purpose of preparing learners for the English they are likely to encounter in real life than (parts from) dialogues that have actually occurred. In order to introduce the use of present progressive forms in yes/no-questions (which is the purpose of the given dialogue extract), it would possibly make more sense to present pupils with authentic examples of progressives that are in fact, as a detailed large-scale analysis has shown (cf. Römer in preparation), comparatively common in interrogative contexts. The forms *happening*, *talking*, *listening*, and *staying*, for instance, occur particularly often in questions. Hence I would argue that the examples in (2) to (5), all from the spoken part of the BNC, may be a better choice than the *playing*- and *packing*-examples in (1) (the phrase "are they packing" does not occur in BNC-spoken; "are they playing" occurs only once).

- (2) What's happening now, does anybody know?
- (3) What are we talking about, what's the subject?
- (4) Are you listening to me?
- (5) Are you staying at your mum's tonight? No. I'm staying at Christopher's.

Doubts about the use and usefulness of invented language in ELT have of course been expressed by several distinguished linguists before and many arguments have been put forward in favour of authentic examples. Firth observed that many of the examples found in grammar books (he quotes the sentence "I have not seen your father's pen, but I have read the book of your uncle's gardener.") are "just nonsense" from a semantic point of view (1957: 24). Sinclair in his 1991 seminal monograph calls it an "absurd notion that invented examples can actually represent the language better than real ones" (ibid: 5). One of the advantages of a real example clearly is that it has in fact occurred in real speech or writing and that it is thus part of the "used" language in Brazil's terms (cf. Fox 1987: 143). Sinclair's well-known precept for language teaching "[p]resent real examples only" is followed by the entirely plausible statement "[l]anguage cannot be invented; it can only be captured" (1997: 31). Another reason why we might want to replace invented examples with real ones is that we may "hinder the development of fluency by excluding data samples that fluent native speakers actually say" (de Beaugrande 2001b: 39). The confrontation with larger amounts of authentic language material will probably help learners become more confident in their use of the foreign language and help them achieve a greater degree of naturalness (cf. Fox 1987: 149).

The assumption that "textbooks are more useful when they are based on authentic native English" (Granger 1998: 7) is, however, not shared by all linguists, some of whom do not seem to be very much in favour of authenticity in the classroom or do not consider the use of invented sentences problematic. Cook states that "[t]he utterances in attested data have also been invented, though for communication rather than illustration." (2001: 376) and thus plays down the attested versus invented sentences problem, putting both kinds of examples on one level. A major difference, however, still lies in the "for communication" Cook mentions. Attested utterances "invented" by competent speakers in a communicative context will probably differ significantly from utterances invented by materials designers in order to illustrate a certain language phenomenon. The former, being genuine examples, arise from a specific context and serve a particular pragmatic function while the only purpose of the

latter is the exemplification of the grammatical structure dealt with in the specific textbook unit. Widdowson even considers it "impossible" to use authentic English in a language teaching context and states that "[t]he language cannot be authentic because the classroom cannot provide the contextual conditions for it to be authenticated by the learners." (1998: 711) Even if it may sound cogent that we cannot transfer the whole context of an actual conversation into the classroom, it should be possible and worth trying to transfer at least part of it and thus achieve a higher degree of authenticity in ELT. On a similar note with reference to the problem of contextual transferability, Michael McCarthy has stated that students are used to and usually very good at recontextualising things because they do that all the time, e.g. while watching soap operas on TV.³ Therefore, I would like to claim that there is a need for more authentic, naturally produced, non-invented examples in EFL teaching.

4. GEFL TC: Corpus compilation and composition

Having stressed the need for authenticity in ELT and having said something about the research potential of an EFL textbook corpus, I will now come to a description of one particular corpus, the German English as a Foreign Language Textbook Corpus (GEFL TC). The need to compile this corpus arose as there were no ready-made computerised collections of German EFL textbook texts available and as it would have been a rather time-consuming undertaking to read through several textbook volumes each time I was looking for a particular language item in order to examine its use.⁴ Besides, all the advantages of any computer corpus (as opposed to non-computerised data collections) also hold true for an electronic collection of textbook text. One such advantage is the possibility to examine at a glance many occurrences of a certain word or phrase in context.

According to Hunston's definition (2002: 16) GEFL TC could be classified as a "pedagogic corpus".⁵ It consists of texts taken from twelve volumes of two introductory course book series (six volumes each) widely used in English language teaching in German secondary schools: *Green Line New* and *English G 2000*.⁶ The texts chosen are all supposed to represent spoken language. Exclusively written material, such as narratives, letters, or excerpts from novels, is not included in the corpus. Spoken texts were selected to enable a comparison with the spoken part of the BNC, which is used as the main source of authen-

tic data in our comparative grammatical studies. This subcorpus was in turn selected because the importance of spoken (based) language is very much stressed in the language teaching curriculum. There is a strong call for teaching materials which help to improve the pupils' communicative competence and prepare them for any prospective discourse with native or near-native speakers of English. Spoken texts appear to better serve this purpose than written samples. Examples of spoken-type texts included in GEFL TC are dialogues, interviews, speech bubbles, and narrative texts mainly consisting of dialogue.

Figure 1 gives an account of the composition of the corpus. The two subcorpora (*English G 2000* and *Green Line New*) are of similar size and internal structure and thus offer the possibility of inter-textbook comparisons. The whole corpus counts 108,424 tokens, which is not a very impressive size for a language corpus judged by today's standards. It has to be kept in mind, however, that we are here dealing with a specialised corpus, which is supposed to represent the language of German secondary school level EFL course materials. Although such a small corpus can hardly be labelled representative of school English in general, it can still reveal a lot about the typical features of textbook language. With a corpus of this size certain kinds of analyses (e.g. of lower frequency language items) can of course not be made and the method of approaching the data has to be chosen accordingly (cf. Sinclair 2001).⁷

The corpus was compiled in four steps. First of all, appropriate pages (i.e., pages including spoken-based textual material) from the twelve coursebooks (see appendix 1) had to be selected. Each single page was digitised with the help of a scanner. The resulting image files could then be processed by OCR software. As figure 2 shows, the parts of the text which were to be included in the corpus had to be highlighted manually (see grey shadings). Thus narrative passages and pictorial material could be excluded. The last step in the compilation of GEFL TC involved a conversion of the data into text format. It was then possible to analyse the text files with a concordance program (in our case WordSmith Tools, Scott 1996).

5. "If you eat your hat, you'll be ill" – An example of corpus analysis with GEFL TC⁸

The starting point for the empirical analysis reported on in the following paragraphs was the fact that if-clauses are often described as a grammatical

Composition of GEFL TC

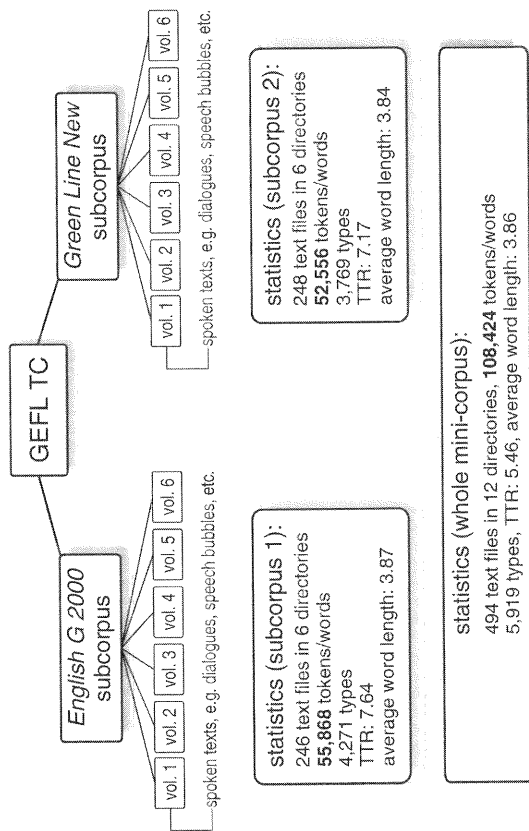


Figure 1. The composition of GEFL TC

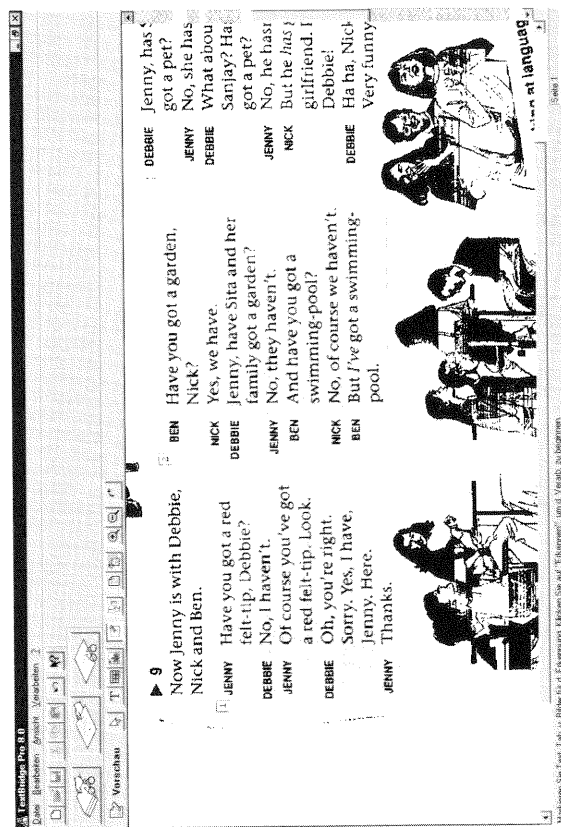


Figure 2. Processing of textbook pages with OCR software (here TextBridge Pro 8.0)

- simple present (if-part) – present tense modal (mainly “will”) + infinitive (main part), as in “Kay will be angry if I come back without his sword.” (Ashford et al. 1997:52)
- simple past (if-part) – past tense modal (mainly “would”) + infinitive (main part), as in “Rodney, if I was worried about stuff like that, it would make me crazy.” (Schwarz 1999b: 89)
- past perfect (if-part) – past tense modal “would” + have + participle (main part), as in “I wouldn’t have been there if I hadn’t gone the wrong way.” (Ashford et al. 1998:37)

In textbooks and school grammars these three combinations are usually referred to as “type 1”, “type 2”, and “type 3” conditionals, respectively. A legitimate question one might ask in this context concerns the didactic significance of these labels. According to Bald (personal communication) the type-labels are rather useless as they do not reveal anything about the semantic functions of the clauses or their syntactic structures. Other sequences of tense forms are underrepresented in the textbook corpus, e.g. the simple present – simple present combination, with 23.7% the most frequent if-clause structure in the analysed BNC_spoken concordance lines, if-clauses with past tense modal + infinitive in both parts (e.g. “I’d be very grateful therefore if er you could put your minds to the options” BNC_spoken), and the structures subsumed under “other combinations” in figure 4. For corpus examples of these and other tense form sequences in if-clauses see Appendix II.

As to the collocational analysis, it could be noticed that “if” has got different collocates immediately to the right (in positions R1 and R2) in BNC_spoken and EFL textbook corpus data. While the personal pronoun “you” is by far the most frequent right-hand collocate of “if” in spoken British English, the cluster “if you” is only one among other frequent clusters in GEFL TC. Combinations that are much more frequent in this corpus than in BNC_spoken are “if I”, “if we”, “if they”, and “if he”. In R2 position the textbook if-clauses feature a number of past participles, especially “was” and “had”, none of which could be found in this position in any of the examined BNC_spoken concordance lines.

6. Pedagogical implications: What can be done to improve teaching materials?

As we have seen in this short analysis, there are some significant differences between the ways if-clauses are presented in GEFL TC and in the spoken part of the British National Corpus. These differences do not only apply to if-clauses taken from the textbook volumes for beginners (volumes 1–3) in which we would probably expect and approve of a simplified, less complex kind of language, but also to examples from some of the last units in the courses (textbook volumes 5 and 6) aimed at more advanced learners.

A closer look at the type of English which pupils are confronted with in the EFL classroom helps us discover what exactly needs to be changed if we want to develop more authentic teaching materials, i.e. if we want to bring the linguistic features we teach closer to the linguistic features we observe in use. The use of examples from (spoken) corpora instead of invented or constructed sentences like “If you eat your hat, you’ll be ill” may be a step in the right direction. Lexico-grammatical items (e.g. different types of if-clauses) could be presented in roughly the same proportions as used in “real” English. We are not suggesting here that one should take our empirical findings to the limit and use exactly the same distribution of if-clause types that were found in spoken corpus data also in EFL textbooks. It would probably mean asking too much of the pupils were they confronted with this huge variety of naturally-occurring tense form combinations. However, we ought not to conceal from learners the fact that if-clause types 1, 2, and 3 are not the only possible (and grammatical) types. There probably is not much use in teaching pupils things about a language which we know are not typical of real language use. Learning a variety of English they will rarely encounter in real-life situations is very unlikely to help learners communicate successfully with competent speakers of English. In this context Gilsan and Drescher claim quite convincingly that “[...] if grammar is to be taught for communicative purposes, the structures presented should reflect their use in current-day native speaker discourse” (1993: 24).

Also, it might be worth paying more attention to collocational patterns and contextual phenomena that are found in native speaker corpus data. A higher degree of authenticity can be achieved if lexical items (such as “if”) are presented in the right context, i.e. the kind of context in which they typically appear in actual language use. As a final pedagogical implication, we could try to make language learners less afraid of using a structure for which there is no

example in their textbooks. An if-clause that can neither be categorised “type 1”, nor “type 2”, nor “type 3” is not necessarily an ungrammatical or unacceptable if-clause.¹¹ A quote from a contemporary English novel illustrates this point quite nicely. In this novel an English teacher tells one of his foreign students to try to get over her hang-up about speaking English. He says: “Don’t let it become too important, okay? ... It doesn’t matter if it comes out sounding different from the textbooks...” (Parsons 2001: 164). In the light of what has been described in the present paper we might even want to add something like “... it is sometimes even more natural if it does not”.

7. Conclusion

The mismatches between BNC_spoken and GEFL TC data make it clear that, at least with respect to if-clauses, the language of German EFL textbooks does not mirror authentic language use. Thus, further studies on lexico-grammatical phenomena in textbook corpora may show how an improvement of English language teaching materials can be achieved on the basis of native speaker corpus data. More analytical and comparative studies of textbook language (and ideally of classroom language as a whole) will be necessary to discover more about the kind of English we teach, its differences from “real” English, and the status of authenticity in ELT. I agree with Glisan and Drescher, who state that “authentic language must continue to be examined if we are to use *real* language as the basis for our teaching” (1993:32), and I am convinced that further corpus-informed comparisons of authentic English and “school” English may lead to fruitful insights for linguists, language teachers, and language learners.

Acknowledgements

I acknowledge the support of my supervisor Wolf-Dietrich Bald and of Walter Pape, Dean of the Arts Faculty, University of Cologne, who helped me fund my participation in the TaLC5 conference in Bertinoro. I would also like to thank David Oakey for helpful comments on an earlier version of this paper and my TaLC audience for stimulating questions and remarks after my presentation.

Notes

1. It has to be kept in mind that the language in an EFL textbook does of course not represent classroom language in its entirety. It constitutes, however, a considerable part of school English, especially if we take into account that ELT (at least in German secondary schools) is very much based on textbooks.
2. For a detailed discussion on different types of multi-word items and for information on lexical connections between words in general see Moon 1997.
3. The statement was made during a discussion at the first Inter-Varietal Applied Corpus Studies Group (IVACS) Conference in Limerick, Ireland, June 15th 2002.
4. The major part of (if not all) empirical analyses of German EFL textbook language so far were carried out manually. Worth mentioning in this context are the works of Dieter Mindt and his colleagues at the *Freie Universität Berlin* (cf. e.g. Mindt 1987, Haase 1995, Schlüter 2002).
5. Another corpus of this type is the TEFL Corpus assembled by the COBUILD team in the mid-1980s. For a description of this project see Renouf 1987.
6. Pupils who use these textbooks are usually about ten years old at the beginning of the course and about sixteen at the end of it.
7. According to Sinclair the difference between corpora is one of method rather than one of size: “There is thus a fairly sharp contrast in method; the so-called small corpora are those designed for early human intervention (EHI) while the large corpora are designed for late or delayed human intervention (DHI)” (2001: xi).
8. The sentence quoted in this headline is typical of the kind of example if-clauses that can be found in the coursebooks included in the analysis. The example is taken from *English G 2000*, vol. 3 (Schwarz 1999a: 29).
9. On this topic see also Bald (1988), who lists conditionals as one of the core problems in English grammar, not only for German learners but for learners in general.
10. Thanks go to Sven Naujokat for providing me with these examples from one of his grade 12 pupils’ in-class essays (grade 12 is the penultimate year before the A-levels).
11. This is not supposed to imply that textbooks exclusively talk about if-clause types 1–3. In textbook and grammar sections for more advanced learners there is some information on mixed conditionals, usually a combination of the type 2 and type 3 if-clauses, 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” (Schwarz 2001: 63) and on so-called “zero conditionals” (simple present in both parts) used to state general validities, e.g. “If you’ve got something with “Made in Sheffield” on it, that’s quality.” (Schwarz 2001: 31). The main focus, however, is on the three types mentioned above.

References

- Amor, S. 2002. *Authenticity and Authentication in Language Learning: Distinctions, Orientations, Implications*. Frankfurt: Peter Lang.
- Bald, W. D. 1988. If-Sätze im Englischen. In: Bald, W. D. (ed). *Kernprobleme der englischen Grammatik* (pp. 38–50). Berlin: Langenscheidt-Longman.
- Beaugrande, R. de. 2001a. "If I were you...": Language standards and corpus data in EFL". *Revista Brasileira de Linguística aplicada* 1 (1): 117–154. online: <http://beaugrande.bizland.com/Ifwereyou.htm> (visited: 20.12.2003)
- Beaugrande, R. de. 2001b. Twenty challenges to corpus research. And how to answer them. online: <http://beaugrande.bizland.com/Twenty%20questions%20about%20corpus%20research.htm> (visited: 20.12.2003)
- Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Breen, M. P. 1985. "Authenticity in the language classroom". *Applied Linguistics* 6: 60–70.
- Conrad, S. 2000. "Will corpus linguistics revolutionize grammar teaching in the 21st century?" *TESOL Quarterly* 34: 548–560.
- Cook, G. 2001. "The philosopher pulled the lower jaw of the hen.' Ludicrous invented sentences in language teaching". *Applied Linguistics* 22 (3): 366–387.
- Firth, J. R. 1957. *Papers in Linguistics 1934–1951*. London: Oxford University Press.
- Fox, G. 1987. "The case for examples". In *Looking up*, J. McH. Sinclair (ed.), 137–149. London: Harper Collins.
- Glisan, E. and V. Drescher. 1993. "Textbook grammar: Does it reflect native speaker speech?" *The Modern Language Journal* 77 (1): 23–33.
- Granger, S. (ed) 1998. *Learner English on Computer*. London: Longman.
- Haase, I. 1995. *Konditionalsätze im authentischen Sprachgebrauch und im Englischunterricht. Eine empirische Untersuchung*. Berlin: Privatdruck.
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Mindt, D. 1987. *Sprache – Grammatik – Unterrichtsgrammatik. Futurischer Zeitbezug im Englischen I*. Frankfurt: Diesterweg.
- Mindt, D. 1996. "English corpus linguistics and the foreign language teaching syllabus". In *Using Corpora for Language Research*, J. Thomas and M. Short (eds), 232–247. London: Longman.
- Moon, R. 1997. "Vocabulary connections: Multi-word items in English". In *Vocabulary: Description, Acquisition and Pedagogy*, N. Schmitt and M. McCarthy (eds), 40–63. Cambridge: Cambridge University Press.
- Parsons, T. 2001. *One for my Baby*. London: Harper Collins.
- Renouf, A. 1987. "Corpus development". In *Looking up*, J. McH. Sinclair (ed.), 1–40. London: Harper Collins.
- Römer, U. (in preparation). *Progressives, Patterns, Pedagogy: A Corpus-driven Approach to Progressive Forms, their Functions, Contexts, and Didactics* (working title).
- Schlüter, N. 2002. *Present Perfect: eine korpuslinguistische Analyse des englischen Perfekts mit Vermittlungsvorschlägen für den Sprachunterricht*. Tübingen: Narr.
- Scott, M. 1996. *WordSmith Tools*. Oxford: Oxford University Press.
- Sinclair, J. McH. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. McH. 1997. "Corpus evidence in language description". In *Teaching and Language Corpora*, A. Wichmann, S. Fligelstone, T. McEnery and G. Knowles (eds), 27–39. London: Longman.
- Sinclair, J. McH. 2001. "Preface". In *Small Corpus Studies and ELT: Theory and Practice*, M. Ghadessy, A. Henry and R. L. Roseberry (eds), vii–xv. Amsterdam: John Benjamins.
- Stubbs, M. 2001. "Texts, corpora, and problems of interpretation: A response to Widdowson". *Applied Linguistics* 22 (2): 149–172.
- Taylor, D. 1994. "Inauthentic authenticity or authentic inauthenticity?" *TESOL-EJ* 1 (2): 1–12. online: <http://www-writing.berkeley.edu/TESOL-EJ/ej02/a.1.html> (visited: 20.12.2003)
- Widdowson, H. G. 1998. "Context, community, and authentic language". *TESOL Quarterly* 32 (4): 705–716.
- Widdowson, H. G. 2000. "On the limitations of linguistics applied". *Applied Linguistics* 21 (1): 3–25.

Appendix I

Coursebooks

- Ashford, S. et al. (eds) 1996. *Green Line New 2*. Stuttgart: Klett.
- Ashford, S. et al. (eds) 1997. *Green Line New 3*. Stuttgart: Klett.
- Ashford, S. et al. (eds) 1998. *Green Line New 4*. Stuttgart: Klett.
- Ashford, S. et al. (eds) 1999. *Green Line New 5*. Stuttgart: Klett.
- Ashford, S. et al. (eds) 2000. *Green Line New 6*. Stuttgart: Klett.
- Aston, P. et al. (eds) 1995. *Green Line New 1*. Stuttgart: Klett.
- Schwarz, H. (ed) 1997. *English G 2000 A1*. Berlin: Cornelsen.
- Schwarz, H. (ed) 1998. *English G 2000 A2*. Berlin: Cornelsen.
- Schwarz, H. (ed) 1999a. *English G 2000 A3*. Berlin: Cornelsen.
- Schwarz, H. (ed) 1999b. *English G 2000 A4*. Berlin: Cornelsen.
- Schwarz, H. (ed) 2001. *English G 2000 A5*. Berlin: Cornelsen.
- Schwarz, H. (ed) 2002. *English G 2000 A6*. Berlin: Cornelsen.

Appendix II

Examples of if-clauses from BNC_spoken and GEFL TC (different tense form sequences, cf. figure 3)

SPr – SPr: Simple present – simple present (“TYPE 0”)

BNC_spoken: we usually attend if it’s in the South anywhere (d95)

GEFL TC: If you help us, this is yours. (GLN, vol. 4, p. 26)

SPr – prtMODAL+inf: Simple present – present tense modal + infinitive (“TYPE 1”)

BNC_spoken: You can use more than one word if you want. (f72)

GEFL TC: If you eat your hat, you’ll be ill. (EG 2000, vol.3, p. 29)

patMODAL+inf – patMODAL+inf: Past tense modal + infinitive – past tense modal + infinitive

BNC_spoken: I’d be very grateful therefore if er you could put your minds to the options (f7a)

GEFL TC: If the carpet makers couldn’t sell their carpets, they wouldn’t have any reason to use children as cheap workers. (EG 2000, vol. 6, p. 49)

SPr – patMODAL+inf: Simple present – past tense modal + infinitive

BNC_spoken: if we let them both know then, then erm, we might we might speed things up. (dch)

GEFL TC: If you don’t, there might easily be misunderstandings and tension. (GLN, vol. 6, p. 53)

SPa – SPr: Simple past – simple present

BNC_spoken: So if you started off in complete darkness you rotate until you get complete darkness or the opposite. (f7u)

GEFL TC: –

SPa – patMODAL+inf: Simple past – past tense modal + infinitive (“TYPE 2”)

BNC_spoken: Lightbulbs would give out more light if they were washed every week in soapy water. (d90)

GEFL TC: Would it be all right if I went to London the weekend after next? (EG 2000, vol. 5, p. 121)

PrPr – SPr: Present progressive – simple present

BNC_spoken: if you ‘re using them in industry, ... then that is the real importance of wearing protective gear. (f77)

GEFL TC: –

PrPr – patMODAL+inf: Present progressive – past tense modal + infinitive

BNC_spoken: if we are moving it to Saturday ... we could switch the venue to Havstock Park (f7j)

GEFL TC: –

SPa – SPa: Simple past – simple past

BNC_spoken: I think we still sent them out if anyone wanted them (f7c)

GEFL TC: the law only allowed executions if the prisoner was able to defend himself (EG 2000, vol. 6, p. 34)

SPa – prtMODAL+inf: Simple past – present tense modal + infinitive

BNC_spoken: if they removed the local authorities judging we ‘ll be in a terrible position! (f7v)

GEFL TC: –

SPr – PrPr: Simple present – present progressive

BNC_spoken: if there ‘s nobody to doing that in Edinburgh they ‘re going to slip again unfortunately. (f7c)

GEFL TC: that’s what’s going to happen if you date him again. (EG 2000, vol. 4, p. 74)

SPr – SPa: Simple present – simple past

BNC_spoken: you did n’t consider this a venue of body building to be suitable if it ‘s not erm it ‘s gon na put bums on seats (d91)

GEFL TC: –

SPa – PrPr: Simple past – present progressive

BNC_spoken: if you did the graph one signature, ... what are we going to call pictures? (f7t)

GEFL TC: –

PrPerf – patMODAL+inf: Present perfect – past tense modal + infinitive

BNC_spoken: if somebody has erm you know, by the end of year seven done that, this would be the national curriculum record (f7e)
GEFL TC:

PrPerf – SPPr: Present perfect – simple present

BNC_spoken: If you have n't done that yet do it now (fmc)
GEFL TC: –

PaPerf – patMODAL+have+PP: Past perfect – past tense modal + have + past participle (“TYPE 3”)

BNC_spoken: –
GEFL TC: I wouldn't have been there if I hadn't gone the wrong way. (GLN, vol. 4, p. 37)

Can the L in TaLC stand for literature?

Bernhard Kettemann and Georg Marko
University of Graz, Austria

Many European language departments combine the study of language and the study of literature. Though the two fields are thus institutionally linked in addition to sharing a common interest in language and texts, they are still treated as completely different in their meta-theoretical and methodological conceptions and their practices of textual analysis. More efforts to bridge the gaps are required, bringing the two perspectives of literary studies and linguistics together. This paper, which is a think-piece rather than a research piece, argues that approaching literary texts through corpus analysis may benefit language students on many different levels. We suggest that the use of concordancing can help them in their explorations of texts – prior to or after a first reading. This, in turn, will enhance their language awareness or, to be more precise, their awareness of the contributions of individual linguistics structures to possible interpretations of a literary text. It will further strengthen their discourse awareness, i.e., it will make them see the differences between literary and non-literary texts more clearly.

1. Introduction

The analysis of literary corpora takes a subordinate position in the vast field covered by the conception of TaLC, a position too low-case to be assigned the capital L in the acronym. The reason for this may be that literary language itself is often seen as distinct from “ordinary” language (the paradigmatic position here is Jakobson 1960) and that its analysis by literary studies also appears to be different from the analysis of other types of language in linguistics and its neighbouring disciplines. Investigating literary language with the help of corpus analysis is therefore considered only marginally relevant to language-related teaching outside the realm of literary studies, i.e., in particular to teaching language and to teaching *about* language and its interaction with culture