

Isabella Seeger

MA TEFL/TESL

Module 6

Corpus Linguistics
June – September 2010

Tutor: Gena Bennett

**Exploiting corpora in German EFL contexts:
textbook design, teacher training and discovery learning**

Module 6 – Assignment CL/10/04

How might corpus research be expected to influence your own academic or professional field in the future? You should consider what corpus research has told us about what language is like, and how this knowledge might be exploited in your chosen field.

Word count: 4,382

excluding long quotes, translations,
figures, tables and references

Translations from German (in true italics) by the author

Table of contents

1	Introduction	1
2	Corpora – collections of real language	2
	2.1 The real McCoy: access to authentic English	2
	2.2 The authenticity of corpora	3
	2.3 Different types of corpora	3
	2.4 Concordancing	4
3	Corpus-informed textbooks and reference works	5
	3.1 Standard EFL textbooks in secondary education	5
	3.2 Teaching materials for adult learners	8
	3.3 Dictionaries and student grammars	9
4	Corpus linguistics in teacher training	10
	4.1 Initial teacher training programmes	10
	4.2 Convincing old hands: overcoming technophobia and traditional beliefs	11
	4.3 Teachers as learners	12
5	Discovery learning: learners as researchers	12
	5.1 Filtered concordances for lower-level learners	12
	5.2 Autonomous research by advanced learners	14
	5.3 Learner corpora vs. natural language use	18
6	Conclusion	18
	Reference list	20

1 Introduction

Considering the natural burgeoning of research in corpus linguistics (CL) with the development of electronic databases, it seems strange that, despite the obvious relevance of easily accessible collections of authentic language to language learning, corpora have been slow to find their way into the world of ELT (Sinclair, 2004, O'Keeffe, McCarthy and Carter, 2007, Boulton, 2007). However, since teachers have begun to recognise the applicability of CL to language learning purposes, "corpora are almost part of the pedagogical landscape" (Sinclair, 2004: 2).

As a German EFL practitioner teaching learners with various needs and proficiency levels, I am interested in the actual and potential influences of CL on language learning and teaching in my local context. However, despite recent changes of the German ELT curriculum aiming to address a perceived lack of communicative competence, and notwithstanding a remodelling of teacher training programmes towards European standards, Mukherjee (2004), Braun (2005), Römer (2006), Rühlemann (2008) and Breyer (2009) point out that in German ELT contexts the field of CL remains largely unexploited, "if not *terra incognita* altogether" (Braun, 2005: 48).

In this paper, I will briefly describe corpora and discuss their authenticity in language learning contexts, and I will present three areas of ELT in Germany which may be influenced by CL in the future: textbook design, teacher training and classroom activities. Firstly, I will demonstrate the importance of spoken corpora for textbook design and reference works in terms of authentic language use and communicativeness. Secondly, I will discuss the advantages (and problems) of familiarising non-native speaker trainee teachers and practitioners with corpus work through learner experience. Lastly, by presenting examples of classroom activities tailored to the specific needs of different target groups, I will show how learners, despite certain difficulties, may become aware of real language use through active corpus research in order to enhance their ability of using language in a more natural way.

2 Corpora – collections of real language

Although bodies of naturally occurring language have been collected and analysed by linguists before the advent of computers, the latter allow not only easy access but also the virtually unlimited storage of written, spoken and audio-visual data and secondary information. The first digital corpora in the 1960s, like BROWN and LOB – one-million-word corpora containing various genres of written American and British English respectively – were tiny compared with today's 100-million-word written and spoken British National Corpus (BNC) or the 500-million-word Bank of English (BoE) of written and spoken British and American language, both comprising a number of subcorpora.

For many research purposes, corpus size is relative, depending on the genre and the number and length of the individual texts contained, but in general larger corpora will serve more purposes than smaller ones (Sinclair, 1991, Crystal, 1995). To be truly representative, corpora must be principled, i.e. reflect all language variables or types to be analysed (O'Keeffe, McCarthy and Carter, 2007, Reppen, 2010). For qualitative or quantitative corpus analysis, concordancing software tools (cf. section 2.4) enable the fast selection, computation and presentation of huge amounts of language samples, thus allowing interpreting language and drawing conclusions about it. The following section suggests how accessing real language sources may influence ELT in Germany.

2.1 The real McCoy: access to authentic English

In my teaching environment, a remote area of Germany, encounters with authentic English were rare before the advent of the Internet. However, even the Internet, providing all kinds of written text, film and even live communication, cannot replace extensive exposure to language in its cultural context. Neither can a corpus; but it can instantly provide large numbers of individual language items in sentence context, which it might take a lifetime to encounter naturally (if at all), thus supplementing and rectifying language users' intuition about language (Hunston, 2002, Bernardini, 2004, Mauranen, 2004a). For teachers (mostly non-native speakers) and learners in Germany, therefore, corpora are particularly valuable as an easily accessible stock of preserved authentic

language. The use of the term 'authenticity' in this context being hotly debated among linguists, it will be briefly discussed in the following section.

2.2 The authenticity of corpora in language learning

It has been argued that corpora, although containing authentic language samples, are inauthentic because transposition from one medium to another decontextualises and equalises the language (Widdowson, 2000), so that it lacks communicative intent and the receiver's perspective (Cook, 1998). However, carefully selected real language samples, albeit decontextualised and culturally alien to some readers, are preferable to contrived, and thus even less authentic, materials (O'Keeffe, McCarthy and Carter, 2007), and Mauranen (2004a) points out that the similarity of corpora to original language suffices for language learning purposes. Mishan (2004: 222) distinguishes between authenticity in terms of genuineness of language and authenticity in terms of learners' engagement with language, as for example in data-driven learning (DDL) (cf. section 5), which contains "an aspect of genuine discovery".

This paper, therefore, takes the stance that, due to the element of authenticity inherent in corpora themselves and also in their use in language teaching contexts, corpora constitute an essential resource of authentic language for the communicative classroom; more so as different types of corpora reflect different types of language, of which the following section will present some examples relevant to ELT in Germany.

2.3 Different types of corpora

Sinclair (1991) distinguishes between static sample corpora containing numerous short texts randomly selected within classified genres, and dynamic monitor corpora reflecting language evolution. Of the different more recent types, I will describe three that are particularly relevant in German ELT contexts (as defined by Hunston, 2002, and Bennett, 2010):

- *General (or reference) corpora*: Large sample corpora comprising subcorpora of different genres of written and spoken language, used for reference purposes and enabling generalisations about language.

Examples: BROWN, LOB, BNC, BoE, the American National Corpus (ANC) and the Corpus of Contemporary American English (COCA).

- *Specialised corpora*: Corpora containing texts of a particular genre, used in specific professional settings and often collected by researchers themselves. Examples: the Cambridge and Nottingham Corpus of Discourse in English (CANCODE), the Michigan Corpus of Academic Spoken English (MICASE) and the Michigan Corpus of Upper-level Student Papers (MICUSP).
- *Learner corpora*: Corpora of written and/or spoken language produced by learners, used to investigate typical errors and for classroom activities. Example: the International Corpus of Learner English (ICLE).

Sections 3 and 5 will make suggestions about the applicability of these and other corpora in textbook design and classroom activities, for which a concordancer (a special software tool) is needed, the basic functions of which I will describe below.

2.4 Concordancing

Concordancers are corpus access programmes not unlike Internet search engines, with more complex functions; there are corpus-related concordancers (the Interactive Corpus Access Tool in the BoE, the BNC Sampler, etc.), independent concordancers (MonoConc, WordSmith, etc.), corpus-building concordancers (e.g. TextStat) and sophisticated ones featuring vocabulary profilers and exercise makers (The Compleat Lexical Tutor, CorpusLab, etc.) or enabling corpus searches of the Internet (e.g. WebCorp). Hunston (2002: 39) defines the concordancer basically as

a programme that searches a corpus for a selected word or phrase and presents every instance of that word or phrase in the centre of the computer screen, with the words that come before and after it to the left and right.

With tagged corpora, concordancers can differentiate between parts-of-speech, include word families in the search, etc. In order to investigate keyword ('node') frequency and collocation significance, the nodes or surrounding words can be sorted to make patterns, collocations or different meanings visible and enable observations, interpretations and conclusions about language and

language use. This includes revising researchers' intuitions of 'prototypical' language, often presented in textbooks (Hunston, 2002) but overridden by corpus findings; the following sections will show that this is particularly true for German EFL textbooks vs. spoken corpora.

3 Corpus-informed textbooks and reference works

With regard to German EFL textbooks, being usually ascribed a central role in learners' speaking competence, Vielau (2005: 137) observes:

[D]ie künstliche, unnatürliche Sprache vorkommunikativer Lehrwerk-generationen, in denen nur das zulässig war, was direkt zum primären grammatischen Aufbau des Lernwegs passte, ist daher heute in den Lehrmaterialien kaum mehr anzutreffen (*The artificial, unnatural language of pre-communicative textbook generations permitting only what was perfectly suitable to the primary grammar progression of the learning course is seldom encountered in teaching materials now*).

This view, however, may be wishful thinking, as other researchers deplore a clear mismatch between textbook language and corpus evidence (McCarthy and Carter, 2002, Conrad, 2004, Römer, 2004, Rühlemann, 2008, among others). According to Mauranen (2004a: 96),

[p]edagogical descriptions tend to be far from adequate in L2 teaching materials, and despite ideological lip-service to the priority of speech, spoken language is not often very realistically depicted in textbooks or reference books. [...] Spoken language is therefore a domain where learners need to work out many linguistic features on their own, because they cannot expect enough help from textbooks, teachers, or reference materials.

This is the more alarming as German learners tend to model their English on textbook rather than real language (Römer, 2007). The following sections will point out how the use of spoken corpora might influence textbook language in secondary and adult education to become more authentic and natural.

3.1 Standard EFL textbooks in secondary education

The curricular guidelines for ELT in German secondary education follow the suggestion of the Common European Framework of Reference (CEFR) to promote learners' interactive language skills (Tranter, 2003), aiming to address the lack

of speaking competence, including the phenomenon of "*Abiturspeak*" ('bookish' speech of advanced German A-level students) (Mukherjee, 2004: 247). To comply with these requirements, three standard textbook series (each comprising adapted versions for lower and middle education and grammar schools) have been redesigned; I will discuss the middle-education versions *Red Line* (Hass, 2006), *Camden Town* (Beer et al., 2006) and *English G21(B)* (Schwarz and Biederstaedt, 2006). On the respective websites, *Red Line* and *Camden Town* prominently advertise communicativeness and authenticity:

"alltagsnahes Englisch
(*English reflecting everyday language*)",

"echtes Englisch
(*real English*)"

(Klett Verlag, n.d.),

"authentisch, schülerorientiert, kommunikativ
(*authentic, student-orientated, communicative*)",

"echte Kommunikation i[m] Klassenzimmer
(*real communication in the classroom*)",

"kommunikative Kompetenz von Anfang an
(*communicative competence from the start*)",

"wie im wirklichen Leben
(*as in real life*)",

"lebensnahe Dialoge
(*dialogues reflecting real life*)"

(Diesterweg Verlag, n.d.).

My question to the publishers on which corpus this 'authenticity' was based nevertheless remained unanswered; *English G21*, advertising communicativeness less obtrusively (Cornelsen Verlag, 2010), is based on the BNC (Rimkus, 2010).

How crucial corpus information is to textbook authenticity may be demonstrated by a comparison of the spoken part of the BNC with *Green Line* (Beile et al., 1984) – an earlier grammar school version of *Red Line* –, finding "huge discrepancies between the use of modal auxiliaries in authentic English and in the English taught in German schools" and suggesting changes, especially in the order of introduction (Römer, 2004: 193).

My analysis of the modals in *Red Line*, *Camden Town* and *English G21* shows that, while none of the series has taken up Römer's (2004) suggestion, *English G21* overlaps with Römer's list more than *Red Line* or *Camden Town* do, introducing *would* much earlier, *shall* later. None, however, teaches *will* at a very early point, while all teach *must* (and two, *may*) early, each in contradiction to corpus evidence (cf. table 1).

Vol.	Red Line	Camden Town	English G21 (BNC-based)	Römer (2004)
1	can must	can must	would may must can	will would can
2	may could should will	could should might	could will	could should might
3	shall	shall may	might	must may
4	ought to might	will would	should	shall
5	---	---	---	ought to
6	(in print)	(in print)	(in print)	

Table 1: Teaching order of modal verbs in three standard textbook series, compared with the order suggested by Römer (2004: 195) on the basis of the BNC. (For simplicity's sake, only the full, positive forms are represented.)

Taking the modals as an example of how 'authentic' EFL textbooks disregard real language use, their compliance with CEFR standards of communicativeness must be doubted.

For more authenticity and communicativeness, textbooks should therefore be more informed by spoken corpora – as Mauranen (2004b: 208) emphasises, "[w]e need spoken corpora for teaching the spoken language." While general corpora like the BNC and their spoken subcorpora contain mainly adult speech (cf. fig. 1 below), German secondary students' language use focusses on peer-to-peer communication (e.g. in chatrooms or during school exchanges); therefore I suggest including teenage speech corpora, like the BNC-related Bergen Corpus of London Teenage English (COLT). Teenage speech becoming quickly outdated, building a corpus monitoring international teenage English might be an interesting project.

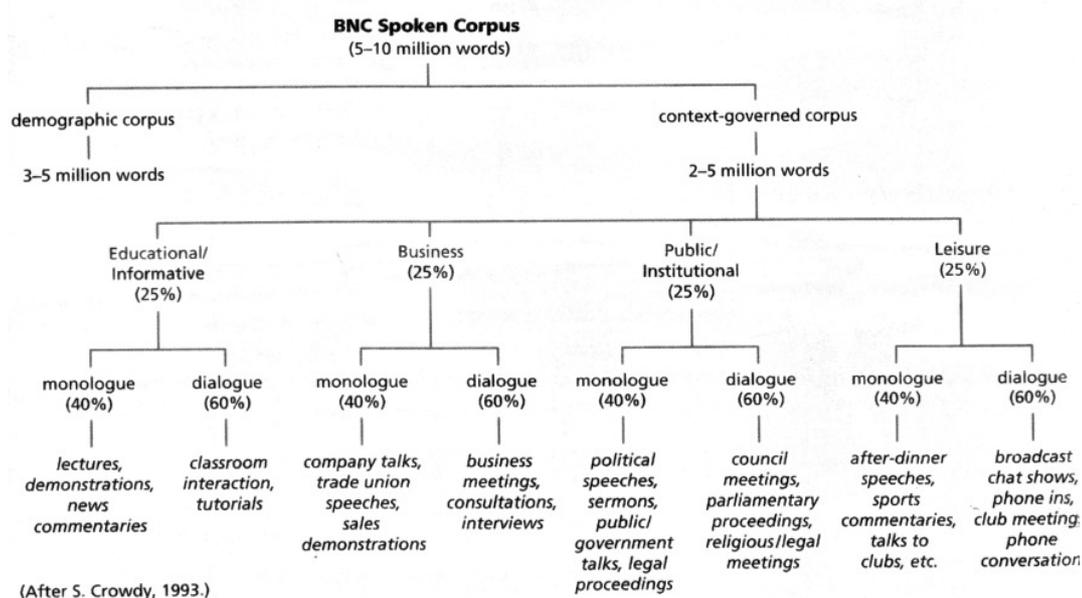


Fig. 1: Components of the spoken part of the British National Corpus. (Crystal, 1995: 440)

3.2 Teaching materials for adult learners

In the absence of a prescriptive ELT curriculum, the variegated needs in adult education have generated a vast range of textbook series and stand-alone materials, treating English as a lingua franca for general or special purposes; because of their abundance and variety I will present only an overall view.

Although many recent materials do feature selected items of conversational speech, a majority of textbooks are still based on Standard English (SE), i.e. written printed language, despite overwhelming evidence for a large discrepancy between SE and spoken corpora (Rühlemann, 2008). Adult teaching materials would therefore also benefit from spoken corpora, as confirmed by Timmis's (2005) study, in which learners and teachers successfully used corpus-informed materials, regarding them useful for communication. Since to my knowledge there exists as yet no explicitly corpus-based German publication, the series *Touchstone* (McCarthy, McCarten and Sandiford, 2005), based on the Cambridge International Corpus of North American English (CICNAE), with a focus on conversational speech (Bennett and Bricker, 2006), offers itself as an alternative. In academic and ESP settings with more specific learners' needs and variegated language types, teaching materials might be informed by available specialised corpora such as MICUSP, MICASE or the Corpus of Spoken Professional American English (CSPA), or by specifically built, smaller corpora of authentic texts from the respective disciplines (Flowerdew, 1996).

3.3 Dictionaries and student grammars

Only a few dictionaries and grammars are standardly used at German schools; in my environment, the bilingual *Langenscheidt Power Dictionary English* (Langenscheidt, 2010) and the *PONS Schülerwörterbuch Englisch für die Schule (Student's dictionary of English for school use)* (PONS, 2010a) are mainly used in years 5-12. Both make no reference to corpus use, the *PONS* limiting itself to "Wortschatz, den man in der Schule wirklich braucht: Wortschatz aller aktuellen Schulbücher, Shakespeare- und Oberstufenwortschatz (*Vocabulary really needed at school: the vocabulary of all current textbooks, Shakespeare vocabulary and vocabulary for years 11-12*)" (PONS, 2010b). This close relation to standard textbooks leads back to the problem discussed in section 3.1. In higher secondary (years 11-12) and tertiary education, the monolingual *Oxford Advanced Learner's Dictionary (OALD)* (Turnbull, 2010) is widely used, most teachers not knowing, however, that it is corpus-based (Mukherjee, 2004), despite its special web feature about the BNC (Oxford University Press, 2010). An alternative is the *Collins COBUILD Advanced Dictionary* (Harper Collins, 2009), a product of the COBUILD project, which also created the BoE.

Student grammars are usually contained in the textbooks (for lower levels) or adapted to textbook series, like *Learning English: Grundgrammatik* (Ungerer et al., 2004), in my environment the standard work for years 11-12, based on the BNC but adapted to *Green Line*, which might again lead back to the textbook problem (cf. section 3.1.) For more advanced students or teachers there are monolingual alternatives, e.g. the corpus-based *Longman Student Grammar of Spoken and Written English* (Biber, Conrad and Leech, 2002), or the BoE-based *Collins COBUILD English Grammar* (Harper Collins, 2005).

Obviously there are many opportunities for corpora, in particular spoken corpora, to influence German EFL teaching materials and reference works in terms of naturalness and communicativeness, while an opening to other, corpus-informed resources might provide alternatives. As the design and choice of teaching materials are heavily influenced by teachers' attitudes and beliefs, the following sections will look at corpus-based instruction in teacher training.

4 Corpus linguistics in teacher training

4.1 Initial teacher training programmes

German university programmes for teachers have recently been redesigned from 'Grund- und Hauptstudium' (*elementary and advanced studies*) with two state examinations to bachelor and master programmes. The new ELT curriculum, however, does not include CL; Mukherjee (2004: 244) observes that

[e]ven today, it is still perfectly possible for each and every student of English language and literature in virtually all 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] [sic] to exploit corpora*.

Six years later, a search across the websites of German universities reveals that only a few offer CL, and then optionally towards the end of the bachelor or in the master programme, requiring knowledge of other European languages besides German and English – thus unlikely to attract many students.

However, being complex and time-consuming, corpus studies should be included in initial teacher training (Rühlemann, 2008), as "[c]areful initiation to corpus skills – or most of all corpus thinking – for the classroom teacher is crucial before corpora can conquer ordinary learners in schools and universities" (Mauranen, 2004a: 100). Breyer (2009) advocates the introduction of CL at an early stage – especially for non-native speakers – to enable trainees to familiarise themselves sufficiently with corpora to achieve teaching competence. I fully agree with Römer (2006) who suggests making CL a compulsory subject, to convert the ELT community from 'right-or-wrong' decisions to an awareness of regularities within language patterns. This "missionary work" (Römer, 2006: 128) should also reach practising teachers; section 4.2 will discuss how to overcome likely problems.

4.2 Convincing old hands: overcoming technophobia and traditional beliefs

Mukherjee's (2004) study about corpus training for German EFL practitioners confirms my personal impression that CL is virtually unknown among this community. University courses being unsuitable for practising teachers, in-house workshops (although a challenge for the education budget) would be more appropriate (Mukherjee, 2004, Mauranen, 2004a). However, with a majority of teachers being older than fifty, sceptical of new ideas and unfamiliar with computers, a novel, computer-based approach is unlikely to be met with enthusiasm. To overcome reluctance, it should be pointed out that corpora can relieve teachers (again, especially non-native speakers) from their perceived responsibility as language experts in the classroom (Bernardini, 2004) and serve as native-speaker reference, e.g. when marking unfamiliar language use in advanced students' essays (Römer, 2006).

Although using corpora and concordancing software requires only basic computer literacy, teachers need to practise enough to become confident about their ability to use the software tools and teach others to use them (Mauranen, 2004a). Computer-literate, CL-trained colleagues (cf. section 4.1) as co-trainers might help to bridge the technological generation gap.

The biggest challenge, however, would be convincing German EFL practitioners to take a new view of language. Very few teachers have ever had extensive experience with real English; to most, language is an "orderly world of clear-cut grammatical rules and clear right-or-wrong decisions" threatened by the intrusion of 'untidy' real language (Römer, 2006: 128). Teachers therefore tend to regard language variation as a nuisance and too complex to be taught (Conrad, 2004, Rühlemann, 2008). However, in the light of corpus evidence and since the aim of ELT in Germany is natural communicative competence (cf. section 3.1), teachers must revise this attitude and realise that natural language features regularities within patterns which learners may discover themselves via corpora. Their own experience as learners enables teachers, in particular non-native speakers, to guide their students in this autonomous discovery, as will be discussed below.

4.3 Teachers as learners

Many non-native speaker teachers lack confidence about their own L2 competence (Árva and Medgyes, 2000), but while their intuitions about language use may indeed be limited, Bernardini (2004: 28) also sees advantages:

Being life-long language learners as well as teachers, they possess an invaluable repertoire of learning strategies and experience of difficulties and successes that students can draw from, whilst their limited intuitions concerning acceptability and appropriateness are less crucial a problem than they used to be.

This insight into the learners' role may also lead to a more democratic teacher-learner relationship, an ideal environment for autonomous, form- and meaning-focussed learning with corpora (Bernardini, 2004). While having to re-evaluate their traditionally dominant role may be challenging for teachers (Johns, 1991), reflecting their own learning process – as language learners and corpus users – and their classroom teaching may engender new theories (Breyer, 2009). Section 5 will show how teachers' newly-gained insight can be applied in the classroom.

5 Discovery learning: learners as researchers

Due to the element of discovery contained in exploring corpora for patterns or rules possibly unknown to learners and teachers, inductive, data-driven learning (DDL) has an invaluable stimulating effect on learners' interest and thus on the language learning process (Johns, 1991). Students can access corpora directly in computer labs, today standard equipment at German institutions; however, concordancing requires not only computer literacy and some language proficiency but also good cognitive skills (cf. section 5.2). Nevertheless, concordance exercises can be tailored to the skills of less sophisticated learners, as suggested in the following section.

5.1 Filtered concordances for lower-level learners

For lower-level learners, concordances can be adapted, without losing their element of authenticity or representativeness, by selecting lines that feature only structures appropriate to learner level. German secondary students, often confusing *interesting* and *interested*, thus might compare these concordances from the British and American spoken corpora of the BoE (cf. fig. 2):

1 Ooh Graham what an interesting point.
 2 This is an interesting track.
 3 That's interesting.
 4 Yeah. well it's interesting.
 5 Right. That's interesting.
 6 Yeah. Yeah. Interesting.
 7 Oh that's quite interesting.
 8 I just thought it was an interesting discussion.

9 I don't think they'd be interested in cork.
 10 That was marathon. I was interested in the marathon then.
 11 We think you're only interested in closing places.
 12 But you just weren't interested enough to follow it up?
 13 He was interested in the bird life.
 14 Somebody like myself who's interested in classic systems.
 15 I'm interested in clouds.
 16 We can do that if you're interested.

Fig. 2: Data from the BoE (Aug 2010), filtered for lower learner level and edited in length (i.e., all text beyond the central clauses was omitted and replaced by capital letters or full stops)

Even lower-level students would probably soon discover that *interesting* usually refers to objects ('What?'), while *interested* refers to people ('Who?') and is often followed by *in* and an object of interest; they might also notice the informal use of *interesting* in some lines. (A follow-up exercise with lines 9-16 would change objects into subjects, with lines 1-8 as models, e.g., *I'm interested in clouds* → *Clouds are interesting* / *Ooh what an interesting cloud.*)

Other variants blank out the node or its collocates, or let learners complete concordance lines from the context (Mishan, 2004). Römer (2006: 125) presents a concordance-based fill-the-gap exercise highlighting the semantic and phraseological differences between *speak* and *talk* (cf. fig. 3):

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

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

Fig. 3: "An example of a possible DDL exercise – 'speak' vs. 'talk'" (Römer, 2006: 125)

Similar exercises are quickly designed by a teacher familiar with basic concordancing, or taken from ready-made, photocopiable materials (Hadley, 2002), as supplementary or stand-alone treatment of particular structures. Since German EFL learners, accustomed to colourful, glossy materials, might be daunted by monotonous-looking concordances, creative teachers will enhance materials; Hadley (2002) suggests highlighting keywords with colours. To avoid overtaxing lower-level learners, I fully agree with Bennett's (2010) advice to ask simple, unambiguous questions, carefully select and edit a small number of lines and let students discover together. More sophisticated learners, however, may welcome independent corpus research, as described below.

5.2 Autonomous research by advanced learners

With little empirical evidence about lower-level learners and DDL, the widespread assumption that it is more appropriate for advanced learners is debatable (Boulton, 2007), and Hadley's (2002) study suggests that DDL may indeed benefit less sophisticated, adult students. However, other researchers (Mauranen, 2004a, Bennett, 2010) agree that corpus analysis, requiring certain intellectual capacities, is more suitable for highly proficient, mature learners; Mauranen (2004a: 99) points out that

[e]ven if we wish to be maximally learner-centred, or construct the learner as a 'researcher', he or she needs skills and guidance in dealing with the kind of data a corpus provides. Noticing things in a corpus is an acquired skill even for linguistically relatively sophisticated learners like L2 majors in university departments [...].

This section therefore refers to advanced students of English for academic or professional purposes, focussing on academic writing and presentation (for literature or translation studies, economy, engineering, etc.) or on upper-level interactive skills (as needed in business, aviation, etc.).

Although teacher-indoctrinated German students might resent forsaking their 'tidy' language image, with no authority to consult about the 'new' language, the intellectual challenge of concordancing and the obvious relevance of authentic samples from their particular discipline may nevertheless motivate students to actively explore them. If corpus evidence then yields no answer or raises new questions, teachers may fear a loss of control (Hunston, 2002); in my

experience, however, admitting to being out of one's depth usually increases credibility with students, especially with top-flight professionals feeling reduced to apprentices in the language classroom.

In the following, I will present some examples for recognising patterns, meanings and collocation. For presentations, students might analyse concordances from an academic corpus featuring signal words (e.g. *so*, *then* and *though*), to recognise patterns and draw conclusions (i.e., that *so*, connecting two sentences, leads from one idea to another [cf. figs. 4a-b below]; that *then*, following conjunctions, adds items or events; and that *though*, set off by commas and following noun or verb phrases, contrasts ideas); the activity could be followed by gap-fill exercises (Bennett, 2010).

- | | |
|---|---|
| 1 | <i>so spend some time thinking about the midterm.</i> |
| 2 | <i>so, the economic changes that have brought more and more women into the work force in the twentieth century have clashed with the inability of our society to deal with large numbers of women in the labor force.</i> |
| 3 | <i>so the reduction for the need of women's labor occurs among young, unmarried women in the home who formed some of the early factory workers in the early stages of industrialization in New England.</i> |
| 4 | <i>so it no longer is considered work.</i> |

Fig. 4a: Discovering patterns: *so* followed by main clauses. Extract from Bennett (2010: 62)

- | | |
|---|--|
| 1 | <i>This is to give you a chance to work on the <u>midterm</u>. so</i> |
| 2 | <i>women who work hard and earn less bring less home to their families. And so</i> |
| 3 | <i>Remember, surplus people are producing for surplus, and so</i> |
| 4 | <i>Gradually the notion of women's work sort of gets erased. It's not paid labor in the home, and so</i> |

Fig. 4b: Discovering patterns: *so* preceded by main clauses and links. Extract from Bennett (2010: 64)

By sorting concordances from a general corpus according to patterns and content, highly proficient students, e.g. translators, may discover various L1 meanings of an L2 word (e.g. *deny*) and their respective significance. Patterns like two objects, *confirm or, whether or that* + main clause and the nature of the objects provide clues to the different meanings of *deny*, and students may conclude that its most significant meanings are *refuse* and *contradict* (cf. fig. 5).

<p>whose religions are being used to deny with welfare "reform" that would deny the worst sexism of all faiths, who deny do they stem from? But they mustn't deny the growing trend in the region to deny passed last November which would deny says the government conspired to deny approach, which, in general, is to deny communities. Victorians willing to deny in London, says: "we don't want to deny</p> <p>refuse sth., esp. a right or an enjoyment, to sb. → 'verweigern, vorenthalten'</p>	<p>deny <u>them their fundamental human rights.</u> deny <u>assistance to children born while a</u> deny <u>educational equality to their women,</u> deny <u>us our own culture.</u> People would suffer deny <u>first asylum</u> was threatening the deny <u>basic public services to illegal</u> deny <u>their client a fair trial</u> by tapping deny <u>retrial of an acquitted person</u> but to deny <u>themselves the satisfaction of what</u> deny <u>young people pleasure,</u> but drugs are</p>
<p>tennis star refuses to <u>confirm or</u> deny said: "we never <u>confirm nor</u> deny PM's office refused to <u>confirm or</u> deny forces are battling Azerbaijan, deny - and the Government does not deny of the community, but city officials deny have animal souls. Oh, I don't deny like "hell" and "bishop", some deny has we brought them to? You couldn't deny</p> <p>contradict a statement or an opinion → 'bestreiten, dementieren'</p>	<p>deny <u>whether she and Sergei are still an</u> deny <u>whether particular tax changes are</u> deny <u>the claim</u> last night. But Mr Prescott deny <u>the Azeri report.</u> Armenian forces took deny <u>it knew</u> that Prof Littlechild was deny <u>that a truce even exists.</u> NPR's Jacki deny <u>that every so often some bird thinks</u> deny <u>that Chaucer had any genuine sense of</u> deny <u>that the house was as sense-assaulting</u></p>
<p>for tests. <p> The Wrights deny Mr Javed's leg. <p> Baker and Power deny can <u>kick an opponent in the face,</u> deny</p> <p>disavow an offence → 'abstreiten'</p>	<p>deny <u>manslaughter and cruelty.</u> <p> Farm deny <u>murder</u> -and along with a teen they deny deny <u>it</u> and still get to play in the FA Cup</p>
<p>fuelled by our prior commitment to deny not eliminate man from my work, deny whether in so doing they choose to deny</p> <p>renounce a belief or conviction → '(ver)leugnen'</p>	<p>deny <u>Christianity</u> in particular, so that deny <u>his existence,</u> and content myself with deny <u>the world of form</u> or affirm its every</p>
<p>and troublesome for your boss to deny suspensions, the Merc would probably deny</p> <p>reject a request → 'ablehnen'</p>	<p>deny <u>you at least some of your demands.</u> If deny <u>their application,</u> according to the</p>

Fig. 5: Sorted concordance lines for deny, with patterns and context clues underlined
(Data from the BoE, Aug 2010)

To rule-adhering German students, collocations are particularly perplexing because "this is simply the way we say things in English and that's that" (Woolard, 2000, cited in Walker, 2009: [III]1). Instead of rules, students can establish and understand 'appropriateness' by measuring collocation frequency and significance in 'Mutual Information (MI) score' or 't-score'. Calculating the strength of word association based on their independent relative frequency, MI-score can be misleading, as infrequent words may score high in particular combinations (e.g., *baleful+gaze*); t-score therefore also includes overall concordance frequency, and a t-score of 2 or higher indicates significance (Hunston, 2002).

Using collocations appropriately may be essential in business negotiating or presenting, owing to frequent L1-L2 incongruencies, e.g. *leiten* meaning *run, head, be responsible for* or *be in charge of* (among others), depending on its collocate, e.g. *company* (Walker, 2009). Learner (and non-native speaker

teacher) intuition being unreliable, students need to know the most significant collocations of these verbs to use them appropriately. The top ten t-scores of their collocations in the BoE (cf. tables 2-5) are:

run

node +2	t-score	freq.
risk	34.14	1,187
business	29.62	946
country	28.40	874
company	22.93	601
show	22.09	539
campaign	13.83	215
club	13.50	226
government	13.20	276
place	13.18	239
school	12.21	202

head

node +2	t-score	freq.
team	11.43	144
government	9.34	112
investigation	9.27	88
inquiry	9.06	84
department	8.87	84
group	8.54	88
committee	7.06	55
CIA	6.60	44
commission	6.48	47
company	5.72	49

responsible for

node +2	t-score	freq.
deaths	15.07	228
actions	13.08	172
death	12.16	153
shareholders	10.02	101
loss	9.70	97
attack	9.31	90
policy	9.26	91
development	9.15	88
murder	9.12	85
safety	8.61	76

in charge of

node +2	t-score	freq.
case	8.77	80
policy	8.59	76
affairs	8.44	72
investigation	8.33	70
project	7.99	65
team	7.76	64
country	6.83	51
operations	6.65	45
operation	6.40	42
development	6.26	41

Tables 2-5: The ten most frequent node +2 nominal collocates associated with *run*, *head*, *responsible for* and *in charge of*. Data from the BoE.
Adapted from Walker (2009: [II]2-4)

With *run+company* scoring by far the highest, students should conclude that the common expression for *eine Firma leiten* is *run a company*; the tables also show that almost all nouns are standardly used with only one of the verbs. Similar knowledge could be gained from specialised corpora for diverse professions and consolidated through various matching exercises (Walker, 2009), while typical errors, e.g. those arising from L1-L2 incongruency, may be analysed via learner corpora (cf. section 5.3 below).

5.3 Learner corpora vs. natural language use

Notwithstanding some traditional scepticism towards error-focussing, comparing learner and native-speaker corpora may make learners aware of typically problematic language use. Discovering cross-linguistic interferences or patterns like overuse of words vs. alternatives inspires a range of follow-up activities, e.g. replacing inappropriate expressions with native-speaker ones (Hunston, 2002, Nesselhauf, 2004) and other consciousness-raising, authentic language activities or grammar exercises, as suggested by Bennett (2010). Although learner corpora do not yield information about individual learner characteristics (learning ability, motivation, etc.), many aspects of spontaneously produced learner data (e.g. influences of L1, proficiency, age and sex or learning environment on learner language) can be analysed at once (Nesselhauf, 2004).

For general purposes, Nesselhauf (2004) suggests using ICLE, a corpus of written language from international university students, combined with the similar Louvain Corpus of Native English Essays (LOCNESS), while corpora built from language used by individual learner groups can be tagged for particular error categories typical of those groups (Bennett, 2010). In my opinion, therefore, learner corpora, combined with comparable native-speaker corpora, are valuable resources for classroom activities.

6 Conclusion

In view of the specific problems in German ELT, namely teachers' and learners' tendency to adhere to rules measuring 'correctness' and the resulting inability of learners to use English naturally, the contradiction between curriculum requirements (i.e. communicative language teaching) and what is actually taught evidently needs amending. Corpora, in particular spoken corpora, show that real English is dramatically different from the language taught in German classrooms; in my opinion, therefore, introducing CL would constitute a momentous shift in current beliefs and help to reshape the German ELT landscape indirectly and directly in three main areas.

The first and most obvious area would be the design of textbooks and student grammars, informed by spoken corpora, to establish a basis for the teaching and learning of natural English. The second area would be teacher

training, introducing student teachers to CL early and offering in-house workshops to practitioners – despite financial hurdles and with a special view to computer aversion and partiality for 'orderly' language –, to enable them to achieve teaching competence through their own role as learners, especially as most teachers are non-native speakers. The last and most delicate area would be discovery learning in the classroom, where time constraints, subject complexity and students' and teachers' belief in authority and 'correctness' on the one hand may be counterproductive, but where, on the other hand, the relevance and 'appropriateness' of authentic language may encourage and motivate in particular advanced students to become themselves researchers and explore the richness of natural language through a corpus to be able to use language more appropriately.

Reference List

- Árva, V. and Medgyes, P. (2000) Native and non-native teachers in the classroom. **System** 28/3: 355-372
- Beer, N. et al. (2006) **Camden Town**. Textbook series. Braunschweig: Diesterweg
- Beile, W. et al. (eds.) (1984) **Green Line**. Textbook series. Stuttgart: Klett
- Bennett, G.R. (2010) **Using Corpora in the Language Learning Classroom - Corpus Linguistics for Teachers**. Ann Arbor: University of Michigan Press
- Bennett, G. and Bricker, M. (2006) Book review: Corpus-based Classroom Activities. **Heis News** 25/2 July 2006 [online] [http://www.cambridge.org/elt/touchstone/images/pdf/Corpus-based Classroom.pdf](http://www.cambridge.org/elt/touchstone/images/pdf/Corpus-based%20Classroom.pdf) (Accessed Jan 18, 2010)
- Bernardini, S. (2004) 'Corpora in the classroom: An overview and some reflections on future developments.' In Sinclair, J. (ed.) **How to Use Corpora in Language Teaching**. Amsterdam: Benjamins: 15-36
- Biber, D., Conrad, S. and Leech, G. (2002) **Longman Student Grammar of Spoken and Written English**. Harlow: Pearson Education
- Boulton, A. (2007) But Where's the Proof? The need for empirical evidence for data-driven learning. **Proceedings of the BAAL Conference**, September 6-8, 2007. Edinburgh: British Association for Applied Linguistics: 13-16
- Braun, S. (2005) From pedagogically relevant corpora to authentic language learning contents. **ReCALL** 17/1: 47-64
- Breyer, Y. (2009) Learning and teaching with corpora: reflections by student teachers. **Computer Assisted Language Learning** 22/2: 153-172
- Conrad, S. (2004) 'Corpus linguistics, language variation, and language teaching.' In Sinclair, J. (ed.) (2004) **How to Use Corpora in Language Teaching**. Amsterdam: Benjamins: 67-85
- Cook, G. (1998) The uses of reality: a reply to Ronald Carter. **ELT Journal** 52/1: 57-63
- Cornelsen Verlag (2010) **English G21: Konzept (Concept)**. [online] <http://www.cornelsen.de/lehrkraefte/reihe/r-4647/ra-6029/konzept> (Accessed Aug 1, 2010)
- Crystal, D. (1995) **The Cambridge Encyclopedia of the English Language**. Cambridge: Cambridge University Press
- Diesterweg Verlag (n.d.) **Camden Town: Unterrichtswerk für Realschulen und verwandte Schulformen (Coursebook series for Realschule and related school forms)**. [online] <http://www.diesterweg.de/suche/reihenansicht.xtp?id=CAMTOWRS05&allText=1> and http://www.diesterweg.de/realschule/englisch/camden_town_rs/camden_town_rs_info.xtp (Accessed Aug 1, 2010)

- Flowerdew, J. (1996) 'Concordancing in Language Learning.' In Pennington, M.C. (ed.) **The Power of CALL**. Houston: Athelstan: 97-113
- Hadley, G. (2002) An Introduction to Data-Driven Learning. **RELC Journal** 33/2: 99-124
- Harper Collins (ed.) (2005) **Collins COBUILD English Grammar**. 2nd ed. Glasgow: Harper Collins
- Harper Collins (ed.) (2009) **Collins COBUILD Advanced Dictionary**. Glasgow: Harper Collins / Boston: Heinle
- Hass, F. (ed.) (2006) **Red Line**. Textbook series. Stuttgart: Klett
- Hunston, S. (2002) **Corpora in Applied Linguistics**. Cambridge/New York: Cambridge University Press
- Johns, T. (1991) 'Should you be persuaded – Two samples of data-driven learning materials.' In Johns, T. and King, P. (eds.) **Classroom Concordancing**. ELR Journal 4: 1-16
- Klett Verlag (n.d.) **Red Line: Konzeption (Concept)**. [online] <http://www.klett.de/sixcms/list.php?page=titelfamilie&titelfamilie=Red+Line&modul=konzoption> (Accessed Aug 1, 2010)
- Langenscheidt (2010) **Langenscheidt Power Dictionary Englisch**. 2nd ed. Berlin/München: Langenscheidt
- Mauranen, A. (2004a) 'Spoken corpus for an ordinary learner.' In Sinclair, J. (ed.) **How to Use Corpora in Language Teaching**. Amsterdam: Benjamins: 89-105
- Mauranen, A. (2004b) 'Speech corpora in the classroom.' In Aston, G., Bernardini, S. and Stewart, D. (eds.) **Corpora and Language Learners**. Amsterdam: Benjamins: 195-211
- McCarthy, M. and Carter, R. (2002) 'Ten Criteria for Spoken Grammar'. In Hinkel, E. and Fotos, S. (eds.) **New Perspectives on Grammar Teaching in Second Language Classrooms**. New York: Routledge: 51-75
- McCarthy, M., McCarten, J. and Sandiford, H. (2005) **Touchstone**. Textbook series. Cambridge/New York: Cambridge University Press
- Mishan, F. (2004). Authenticating corpora for language learning: a problem and its resolution. **ELT Journal** 58/3: 219-227
- Mukherjee, J. (2004) 'Bridging the Gap between Applied Corpus Linguistics and the Reality of English Language Teaching in Germany.' In Connor, U. and Upton, T.A. (eds.) **Applied Corpus Linguistics. A Multidimensional Perspective**. Amsterdam: Rodopi: 239-250
- Nesselhauf, N. (2004) 'Learner corpora and their potential for language learning.' In Sinclair, J. (ed.) **How to Use Corpora in Language Teaching**. Amsterdam: Benjamins: 125-152

- O'Keeffe, A., McCarthy, M. and Carter, R. (2007) **From Corpus to Classroom: language use and language teaching**. Cambridge/New York: Cambridge University Press
- Oxford University Press (2010) **The British National Corpus: facts and figures**. [online] http://www.oup.com/elt/catalogue/teachersites/oald7/more_on_dicts/bnc?cc=global (Accessed Aug 2, 2010)
- PONS (2010a) **PONS Schülerwörterbuch Englisch für die Schule (PONS Student's dictionary of English for school use)**. Stuttgart: PONS
- PONS (2010b) **PONS Schülerwörterbuch Englisch für die Schule (PONS Student's dictionary of English for school use)**. [online] http://www.pons.de/englisch/woerterbuch/schule_studium/3-12-517099-0 (Accessed Aug 2, 2010)
- Reppen, R. (2010) **Using Corpora in the Language Classroom**. Cambridge/New York: Cambridge University Press
- Rimkus, G. (grit.rimkus@cornelsen.de) (June 2, 2010) **AW: AW: Umfrage (Re: Re: Survey)**. E-mail to [ID 987961] [e-mail address]
- Römer, U. (2004) 'A corpus-driven approach to modal auxiliaries and their didactics'. In Sinclair, J. (ed.) **How to Use Corpora in Language Teaching**. Amsterdam: Benjamins: 185-199
- Römer, U. (2006) Pedagogical Applications of Corpora: Some Reflections on the Current Scope and a Wish List for Future Developments. **Zeitschrift für Anglistik und Amerikanistik** 54/2: 121-134
- Römer, U. (2007) 'Learner Language and the Norms in Native Corpora and EFL Teaching Materials: A Case Study of English Conditionals.' In Volk-Birke, S. and Lippert, J. (eds.) **Anglistentag (Anglistics Day) 2006 Halle**. Proceedings. Trier: Wissenschaftlicher Verlag Trier: 355-363
- Rühlemann, C. (2008) A Register Approach to Teaching Conversation: Farewell to Standard English? **Applied Linguistics** 29/4: 672-693
- Schwarz, H. and Biederstaedt, W. (eds.) (2006) **English G21**. Textbook series. Berlin: Cornelsen
- Sinclair, J. (1991) **Corpus, Concordance, Collocation**. Oxford: Oxford University Press
- Sinclair, J. (ed.) (2004) **How to Use Corpora in Language Teaching**. Amsterdam: Benjamins
- Timmis, I. (2005) Towards a framework for teaching spoken grammar. **ELT Journal** 59/2: 117-125
- Tranter, G. (2003) Der Gemeinsame Europäische Referenzrahmen – Warum? Wie? Was? (*The Common European Framework of Reference – Why? How? What?*). **Klett VHS Tipps, Sprachenservice Erwachsenenbildung** 36/2003: 3-6

- Turnbull, J. (ed.) (2010) **Oxford Advanced Learner's Dictionary**. 8th ed. Oxford: Oxford University Press
- Ungerer, F. et al. (2004) **Learning English: Grundgrammatik (*Elementary grammar*)**. Stuttgart: Klett
- Vielau, A. (2005) 'Lehrwerk, quo vadis? – Einflüsse auf die Lehrwerksentwicklung (*Coursebook, quo vadis? – Influences on coursebook development*)'. In Burwitz-Melzer, E. and Solmecke, G. (eds.) **Niemals zu früh und selten zu spät: Fremdsprachenunterricht in Schule und Erwachsenenbildung (*Never too early and seldom too late: Foreign language teaching in secondary and adult education*)**. Berlin: Cornelsen: 137-147
- Walker, C.P. (2009) Teaching collocation. **Workshop at the CELS Summer School**, July 8, 2009. Handouts I-IV. Birmingham: University of Birmingham
- Widdowson, H.G. (2000) On the limitations of linguistics applied. **Applied Linguistics** 21/1: 3-25
- Woolard, G. (2000), cited in Walker, C.P. (2009: [III]1) Teaching collocation. **Workshop at the CELS Summer School**, July 8, 2009. Handouts I-IV. Birmingham: University of Birmingham