

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/303701018>

Towards corpus literacy in foreign language teacher education: Using corpora to examine the variability of reporting verbs in English

Chapter · January 2016

CITATIONS

16

READS

1,016

1 author:



Marcus Callies

Universität Bremen

79 PUBLICATIONS 380 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Varieties of English in Foreign Language Teacher Education [View project](#)



Language and Gender [View project](#)

Marcus Callies, Universität Bremen

Towards corpus literacy in foreign language teacher education: Using corpora to examine the variability of reporting verbs in English

Abstract

The aim of this chapter is to contribute to research and practice on the integration of corpus linguistics into curricula for foreign language teachers. I discuss the concept of corpus literacy, a bundle of complex skills conceived of as the ability to use the tools and technology of corpus linguistics to investigate language and to enhance foreign language learning and teaching. The chapter introduces some key concepts and categories that pertain to the use of corpora in the foreign language classroom, and in its main part describes a sample activity that makes use of various types of corpora to raise language teachers' and learners' awareness of register differences and aspects of the lexico-grammatical variability of reporting verbs in spoken and written registers of English.

Der Aufsatz versteht sich als anwendungsorientierter Beitrag zur Einbindung korpuslinguistischer Inhalte und Kompetenzen in die fachliche Ausbildung angehender FremdsprachenlehrerInnen. Als theoretische Grundlage dient dabei das Konzept der sogenannten „corpus literacy“, definiert als Bündel komplexer Fertigkeiten zur kompetenten Verwendung korpuslinguistischer Recherche- und Analysetechniken und –werkzeuge, die im Sinne eines datengetriebenen, explorativen Unterrichtsansatzes zur Bereicherung des Fremdsprachenlernens und –lehrens eingesetzt werden können. Der Beitrag führt zunächst in einige zentrale Konzepte und Kategorien ein, die zum Verständnis der Verwendung von Korpora im Fremdsprachenunterricht wichtig sind, und beschreibt dann im Hauptteil eine Übungssequenz, deren Ziel es ist, unter Anwendung verschiedener Korpora die Sprachbewusstheit von LehrerInnen und LernerInnen bzgl. der registerspezifischen Verwendung und lexiko-grammatischen Variabilität sogenannter Berichtsverben im Englischen zu schärfen.

1. Introduction

The compilation and accessibility of computer corpora and software tools for corpus analysis has revolutionized (applied) linguistics in the last decades. The descriptive grammar of English has greatly benefitted from the

corpus-based and corpus-driven investigation of ‘real’ English as it is actually used by speakers and writers in different contexts and for different purposes. Corpora reveal what native speakers typically write or say, and a corpus can thus be considered an operationalized, readily accessible native-speaker informant. Corpus evidence suggests which linguistic structures will most likely be encountered by language users and learners, i.e. what is frequent and typical and thus should arguably deserve more attention in language teaching. Consequently, the advent of corpus linguistics has also impacted foreign/second language (L2) teaching and learning and has led to the development of numerous resources and pedagogical applications in a wide range of areas of English language teaching, for example:

- (learner-oriented) lexicography (corpus-informed monolingual (learner) dictionaries of English such as the “Oxford Advanced Learners Dictionary” or the “Longman Dictionary of Contemporary English”)
- grammar writing (corpus-based, descriptive student grammars of English such as Biber et al.’s (2002) “Longman Student Grammar of Spoken and Written English”)
- teaching materials and syllabus design (“real” examples for textbooks and exercises retrieved from corpora to increase the authenticity of the material)
- classroom methodology (corpora or corpus-derived materials used in a wide range of classroom applications, in particular in Data-Driven Learning (DDL) activities)

In general, however, it appears that the great potential of corpus linguistics for language teaching practice has not yet been fully realized. In a recent survey article on pedagogical corpus applications and publications in the field of corpus linguistics and language teaching, Römer (2011) states that:

[t]he practice of English language teaching (ELT) to date, at least, seems to be only marginally affected by the advances of corpus research, and comparatively few teachers and learners know about the availability of useful resources and get their hands on corpus computers or concordances themselves (ibid.: 206).

Obviously, it is the new generation of language teachers that is of key importance to popularize and implement corpus linguistic tools and methods in the language teaching community at large. To be able to do so, both pre-service teacher trainees and experienced, qualified foreign language

teachers must be provided with what has frequently been referred to in the literature as “corpus literacy”, i.e. the ability to use corpora for language analysis and instruction. Applied linguists have been emphasizing the importance of integrating corpus literacy in teacher training programs, and so have several professional academic associations¹.

The aim of this chapter is to contribute to research and practice on the integration of corpus linguistics into curricula for foreign language teachers. I will first discuss the concept of corpus literacy and its components, and then introduce some key concepts and categories that pertain to the use of corpora in the foreign language classroom. In the main part of the chapter I will describe a sample activity that makes use of various types of corpora to raise language teachers’ and learners’ awareness of register differences and aspects of the lexico-grammatical variability of reporting verbs in spoken and written registers of English.

2. Corpus literacy for foreign language teachers

In 2001/02, Mukherjee (2004) conducted a survey in the context of teacher training workshops on corpus linguistics for qualified English language teachers in Germany. The findings of this survey revealed a huge gap between corpus linguists’ enthusiasm about the language-pedagogical potential of corpus resources and tools on the one hand, and the reality of English language teaching in Germany on the other. At the time of the survey, almost 80 % of the teachers had never heard about corpus linguistics prior to the workshop. After the workshop, the large majority (over 95 %) agreed that English language teaching may profit from the introduction of corpora and corpus tools. However, this majority remained somewhat sceptical about the application of corpora in practice as they clearly favoured

1 See for example the position paper published by the German Association of University Teachers of English (“*Deutscher Anglistenverband*”) and the German Association of American Studies (“*Deutsche Gesellschaft für Amerikastudien*”) entitled “*Leitlinien zu inhaltlichen Anforderungen für Fachwissenschaft und Fachdidaktik in der Lehrerinnen- und Lehrerbildung im Studienfach Englisch (Lehramt Gymnasium, Sekundarstufe II)*”, in which corpus literacy is mentioned prominently (*Deutscher Anglistenverband und Deutsche Gesellschaft für Amerikastudien* 2009).

making use of corpus data and corpus-based methods themselves (i.e. in teacher-centred activities), while giving learners direct access to corpus data by means of learner-centred activities was disfavoured (only 12 % of the teachers considered implementing these in their own teaching). In view of these results, Mukherjee advocated a concerted effort to popularise the language-pedagogical potential of corpus linguistics and to train teachers to acquire corpus literacy. Mukherjee emphasized that first and foremost it is the teachers who needed to be trained and convinced of the usefulness of corpus data to face challenges in their own teaching before they can be expected to implement more advanced, learner-autonomous activities.

Römer (2009) carried out a teachers' needs analysis to gain information about the situation of teachers, their problems and wishes. Römer's questionnaire-based survey examined topics such as the quality of existing teaching materials, authenticity in language teaching, and the teachers' language competence and exam marking. The participants' answers resulted in the following wish list:

- a wider range of better teaching materials with more interesting, longer and genuine texts, and with more exercises
- textbooks that reflect actual language use and cover variation
- textbooks that keep track of language development
- tools which enable an ad-hoc creation of suitable materials, e.g. worksheets on particular lexical items or structures
- help with exam marking
- more reliable and more quickly accessible resources or reference tools that help them solve the question "Can you say that in English?", perhaps "an online service or database", as one teacher wrote
- an always available native-speaker consultant (Römer 2009: 89).

It appears that many of these desiderata could be met, at least partially, by providing teachers with corpus literacy.

Corpus literacy can be conceived of as a bundle of complex skills. Heather and Helt (2012) define corpus literacy as "the ability to use the technology of corpus linguistics to investigate language and enhance the language development of students" (2012: 417). Several subcomponents of corpus literacy have been identified (see Mukherjee 2006: 14, Dalton-Puffer 2014):

- 1) understanding basic concepts in corpus linguistics (What is a corpus and what types of corpora are available and how? What can you do — and cannot do — with a corpus?)
- 2) searching corpora and analysing corpus data by means of corpus software tools, e.g. concordancers (What is corpus software and how can it be used to search a corpus? How can corpus output be analysed?)
- 3) interpreting corpus data (How may general trends in language use/change be extrapolated from corpus data?)
- 4) using corpus output to generate teaching material and activities.

At the time when Mukherjee (2004) carried out his survey², only few university curricula included an obligatory seminar or even module in corpus linguistics. In view of the explosive growth of the field since the early 1990s, it is likely that the situation has changed for the better in that students are increasingly being taught about corpora and how to exploit them (components 1–3 in the list above). The number of linguistic content classes that focus on corpus linguistics or include substantial sections on corpus linguistics can thus be assumed to have grown since the early 2000s. However, still very few universities seem to have implemented specialized linguistic modules for language teacher students that include seminars on how teachers can use and exploit corpora to teach language (component 4).

As for published research on corpus literacy for language teachers, rather few studies identified the specific skills teachers need to competently and successfully integrate corpus linguistics into their teaching practice, reported on the integration of corpus-linguistic content into curricula for language teachers, or set up standards or guidelines for training teachers in corpus literacy (see Heather / Helt 2012 for review). Even fewer studies empirically examined and evaluated the process of initiating teachers into corpus literacy (e.g. Farr 2008; Breyer 2009; Heather / Helt 2012; Leńko-Szymańska 2014a, 21014b). These studies report first encouraging results as “the corpus literacy training received by (...) pre-service teachers appears to have been effective” (Heather / Helt 2012: 436) and “the students reacted positively to the course and they saw the benefits of corpus-based materi-

2 See Tribble (2015) for a more recent survey. Unfortunately, this paper came to my attention too late to be included in detail into the present chapter.

als and tools in language teaching” (Leńko-Szymańska 2014a: 260). Still, Leńko-Szymańska (2014a) concludes that:

[o]nly extensive exposure to corpora by future teachers coupled with suitable teacher training in the applications of corpora in language education may bring a substantial change in the scope of corpus use in language classrooms in the wide educational context (ibid.: 273).

The present chapter reports on the integration of corpus-linguistic content into the curriculum for English language teachers at the University of Bremen, Germany. Before doing so, however, I will discuss some basic categories related to the use of corpora in foreign language teaching.

3. Using (learner) corpora in the foreign language classroom

There is by now a vast literature on the use of corpora in language education, and this reflects the enthusiasm of (applied) corpus linguists about the pedagogical potential of corpus resources and tools³. Römer (2011: 207) draws a general distinction between direct uses (hands-on for teachers and learners) and indirect applications of corpora (hands-on for researchers and material writers). A crucial concept for direct uses of corpora in language learning is Data Driven Learning (DDL), essentially an inductive, discovery-oriented and research-based approach in which students are exposed to corpus data. Depending on the teacher’s and learners’ degree of skills and autonomy, DDL activities range from activities with high learner autonomy, such as the in-class use of corpora to generate concordances and subsequent exploration of L2 patterns/regularities in student-centred corpus-browsing projects, to activities with low learner autonomy in the form of paper-based concordance exercises based on corpus output prepared by the teacher.

Corpora and corpus linguistic tools and methods are also increasingly used in the study of L2 acquisition, in particular in the field of learner corpus research (LCR), an interdisciplinary and fast-growing field at the crossroads of corpus linguistics, L2 research, and foreign language teach-

3 More recent publications are O’Keeffe / McCarthy 2010, chapters V and VI; Campoy-Cubillo / Bellés-Fortuño / Gea-Valor 2010; Reppen 2010; Frankenberg / Flowerdew / Aston 2011; Kübler 2011; Flowerdew 2012; James / Boulton 2012; Leńko-Szymańska / Boulton 2015.

ing. Learner corpora include large amounts of authentic, continuous and contextualized language data produced by L2 learners and have first and foremost been used as a resource to improve the in-depth description of (mostly advanced) interlanguages and to give L2 learning theories a more solid empirical foundation (alongside with experimental data). Learner corpus data also have great potential for teaching purposes as their analysis can provide a systematic account of the difficulties and needs of specific learner populations (for example German learners of English).

Similar to the distinction between direct and indirect uses of native-language corpora, Granger (2009) discusses learner corpora for immediate vs. delayed pedagogical use. In immediate uses, learner corpora are collected and used by teachers as part of regular classroom activities (so-called local learner corpora), while in delayed uses they are not used directly as teaching and/or learning material by learners who have produced the data, but are collected and used by academics and/or publishers to create corpus-informed teaching materials. Despite Granger's assessment that "learner corpus research has (...) given rise to relatively few concrete pedagogical applications" (ibid.: 14), there is one form of applying learner corpora that seems practical and appealing at the same time. Combined with native-language corpora as positive evidence of language use, learner corpora can be used to provide negative evidence, i.e. common and persistent errors (see Nesselhauf 2004). In that way, learner corpus data used in DDL activities can increase learners' (and future teachers') abilities to notice errors. In the next section, I will describe a DDL-activity based on native-language and learner corpora.

4. Using native-language and learner corpora to examine the variability of reporting verbs in English

Setting the scene

The activity outlined in this section constitutes a two-hour session as part of a one-semester linguistics seminar on learner corpus research. The seminar is regularly offered in the context of the University of Bremen's teacher training program ("Master of Education") for future teachers of English at German primary and secondary schools. The curriculum includes an obligatory linguistics module in which students have to take a seminar loosely called "Key Topics in Linguistics for Teachers of English". This

title is used as a shell for a variety of seminars on topics such as “Learner corpus research”, “Corpus linguistics for the foreign language classroom”, “Teaching varieties of English” and “(Electronic) Dictionaries in the foreign language classroom”. The module description is largely skill- and competence-oriented. Among other things, such as that teacher students will learn about the major bi- and monolingual (electronic) dictionaries, how to use them to correct their pupils’ output, and how to teach them to use such dictionaries appropriately and effectively, the description prominently includes information about corpus literacy:

students [i.e. teacher trainees, MC] acquire corpus-linguistic expertise (“corpus literacy”) in order to use computer corpora in foreign language teaching: 1) they will learn how to use corpora to verify/falsify their hypotheses about language use (corpora as a readily accessible native speaker to judge linguistic structures as to their register adequacy and idiomaticity); 2) they know the opportunities of Data-Driven Learning (DDL) in order to make use of these for research-based learning, to prepare their teaching or to develop their own teaching materials (University of Bremen 2015, my translation).

In the proposed activity, teacher students learn about the use of corpora both from a learner’s and teacher’s perspective. The aim is to familiarize them with a set of corpora and corpus methods so that in the long run they will consider corpora as part of the toolbox for language teachers and learners alike, just like dictionaries and other language learning resources.

Describing the structures under study

The linguistic structures used for the DDL activity are lexical verbs frequently used to report views, opinions, beliefs, facts and findings in written English. In the literature on academic writing, such verbs are often collectively referred to as “reporting verbs”. Other terms frequently used include “research verbs” and “discourse verbs”. They are crucial for reporting content, establishing other authors’ and the writer’s own claims and situating these within published research. Hyland (1999: 344) emphasizes that “they allow the writer to clearly convey the kind of activity reported and to precisely distinguish an attitude to that information, signaling whether the claims are to be taken as accepted or not”. For the present purpose, the term “reporting verb” is used in a wider sense than that found in the literature on academic writing. While Thomas / Hawes (1994: 129) narrowly define reporting as “the attribution

of propositional content to a source outside the author of the article”, it is clear that arguments, facts and findings reported in an argumentative essay or academic paper are not restricted to those cited from other sources, but also include those that stem from the author’s own judgment or research.

Various categories and taxonomies of reporting verbs have been proposed in the literature (e.g. by Thompson / Yiyun 1991 or Hyland 1999). However, using the term is not without problems. It is used rather loosely, and there is a variety of other labels to refer to various (sub-)classes of verbs that partially overlap with reporting verbs, e.g. ‘communication and mental verbs’ (Biber et al. 1999: 362) or “argumentative verbs”. It is thus not easy to draw up a concise, standard list of such verbs. Therefore, working with a finite list of the most commonly used reporting verbs is not ideal, but this approach was deemed more practical when having students take their first steps in (learner) corpus linguistics⁴. Based on a review of corpus-based research on reporting verbs (e.g. Hyland 1999: 349, Bloch 2010: 230), a list of seven lexical verbs frequently used to report others’ facts and findings in argumentative and academic writing was compiled, see (1). For pedagogical purposes explained further below, three more verbs that are not as typical of argumentative/academic writing as those in (1), but which have been examined in previous learner-corpus studies (Neff et al. 2003), were added to this list, see (2).

(1) argue, describe, discuss, propose, report, show, suggest

(2) believe, say, think

Selecting and accessing the corpora

Several corpora representing the language of English native speakers/writers and learners of English as a Foreign Language (EFL) are selected to compare different modes and (sub-)registers of language use. For each (sub-)register, one learner and one native-language corpus is chosen:

4 A corpus-based approach using a rather short list of the most frequent reporting verbs used by native-speaker writers is bound to identify and represent only an incomplete picture of how learners use language, because it neglects a wider range of linguistic means for expressing a particular function. A corpus-driven approach, e.g. based on frequency lists of all lexical verbs used in a corpus (see e.g. Granger / Paquot 2009) is more desirable, but was deemed too time-consuming and complex for the present purpose.

- **spoken language (interviews)** taken from the “Louvain International Database of Spoken English Interlanguage” (LINDSEI; Gilquin / De Cock / Granger 2010) and the “Louvain Corpus of Native English Conversations” (LOCNEC)
- **argumentative writing (student essays)** taken from the “International Corpus of Learner English” (ICLE; Granger et al. 2009) and the “Louvain Corpus of Native English Essays” (LOCNESS)
- **academic writing (unpublished research papers and reports written by novice student writers)** taken from the “Corpus of Academic Learner English” (CALE; Callies / Zaytseva 2013) and the “Michigan Corpus of Upper-level Student Papers” (MICUSP; Römer / O’Donnell 2011).

The LINDSEI, ICLE and CALE contain data produced by university undergraduates in their twenties, mostly in the third or fourth year of studies, whose proficiency level ranges from higher intermediate to advanced, assessed on external criteria, i.e. institutional status. The data contained in the native-speaker control corpora were collected from British grammar school pupils taking their A-levels (LOCNESS), British university undergraduates (LOCNESS and LOCNEC) and US-American university/college students (LOCNESS and MICUSP). For the learner corpora, only the components representing German learners of English are selected to match the native language of the majority of teacher students involved in the seminar the activity is designed for. From the LOCNESS and the MICUSP, subcorpora are compiled so as to keep the register/genre variable constant (i.e. only argumentative essays from the LOCNESS and only linguistics research papers and reports written by English native speakers from the MICUSP⁵).

In addition, and to confront the teacher trainees with data that are closer to their own future teaching reality, a learner corpus that contains written data produced by younger learners at the beginning and intermediate levels is used. The “International Corpus of Crosslinguistic Interlanguage” (ICCI; Tono / Díez-Bedmar 2014) contains written data collected from young learners ranging from primary to secondary school pupils across different countries. These learners were assigned short descriptive and argumentative

5 The ICLE also contains literary and creative writing, while the MICUSP contains a larger variety of text types from several disciplines.

writing tasks (see *ibid.*: 170). For the DDL activity, only the Austrian component (L1 German) is chosen. This component mostly includes descriptive texts (91 %) in which the learners describe themselves, write a postcard to a friend or describe their favourite food. A breakdown of all corpora is shown in Table 1, and the distribution of texts per school grade in the ICCI is detailed in Table 2.

Table 1: Corpora used in the DDL-activity

| corpus | variety | professional status | mode | text type | # texts | ~ total words* |
|-------------|----------|---------------------|---------|---------------------------|---------|----------------|
| LINDSEI.GER | EFL | students | spoken | interviews | 50 | 86,000 |
| LOCNEC | BrE | students | spoken | interviews | 50 | 118,500 |
| ICCL.AUT | EFL | pupils | written | descriptive | 773 | 97,500 |
| ICLE.GER | EFL | students | written | argumentative essays | 299 | 153,500 |
| LOCNESS | BrE+ AmE | students | written | argumentative essays | 290 | 210,000 |
| CALE.GER | EFL | students | written | research papers | 55 | 174,000 |
| MICUSP | AmE | students | written | research papers + reports | 18 | 62,300 |

* All word counts were obtained with the help of “WordSmith Tools 5” (Scott 2008) excluding mark-up enclosed in angular brackets.

Table 2: Breakdown of the Austrian component of the ICCI

| grade | # texts | total words* |
|-------|---------|--------------|
| 5 | 60 | 4,700 |
| 6 | 148 | 10,500 |
| 7 | 167 | 22,800 |
| 8 | 140 | 19,200 |
| 9 | 118 | 19,200 |
| 10 | 71 | 9,700 |
| 11 | 69 | 11,400 |
| total | 773 | 97,500 |

* Word counts were obtained with the help of “WordSmith Tools 5” (Scott 2008).

The corpus files are stored in separate folders on a server in the faculty's computer lab and made available to the students for in-class use. The corpora can then be accessed by means of concordance software such as "WordSmith Tools" (Scott 2008) or "AntConc" (Anthony 2015).

Retrieving and processing the data

Students are provided with the list of reporting verbs and corpora. They gained some basic knowledge about key concepts, tools and methods in corpus linguistics during their BA studies⁶. After a quick introduction to the corpora and the software tools, students are divided into four groups of two to four students with each group examining only one (sub-)register. Groups 1–3 are asked to analyse 1) the LINDSEI.GER / LOCNEC, 2) the ICLE.GER / LOCNESS and 3) the CALE.GER / MICUSP respectively, while the task of group 4 is to examine the distribution of reporting verbs in the ICCL.AUT. Each group is thus asked to search for the ten verbs in the corpora, generate concordances, eyeball and filter these manually, and then count and calculate lemma frequencies by means of spreadsheet software.

This first step alone requires a set of highly complex analytic skills that pertain to component 2 of corpus literacy (searching corpora and analysing corpus data by means of corpus software tools) described in Section 2 above:

- knowing how to carry out wildcard/combined searches in the concordance program to retrieve all inflected forms of a verb in one search;
- sorting and inspecting the concordance lines, thereby weeding out false positives such as noun forms like *TV shows*, *traffic reports*, *rational thought*, and disambiguating potentially polysemous instances (e.g. *argue*₁ 'disagree with someone in words, often in an angry way' vs. *argue*₂ 'state, giving clear reasons, that something is true, should be done')
- understanding the concept and necessity of normalising frequency counts because of unequal corpus sizes and then actually calculating normalised frequencies for the occurrence of each verb in the respective corpus.

6 Students of English at the University of Bremen take an obligatory seminar in linguistic research methods in their second semester. The seminar has a strong focus on corpus linguistics. Moreover, most linguistic content classes include an extended session on corpus linguistics.

The resulting spreadsheet tables with raw and normalised frequency counts are reproduced as Tables 3 and 4.

Table 3: Frequency of occurrence of ten reporting verbs in six corpora (normalised frequencies per thousand words are given in the shaded columns; EFL = English as a Foreign Language; ENL = English as a Native Language)

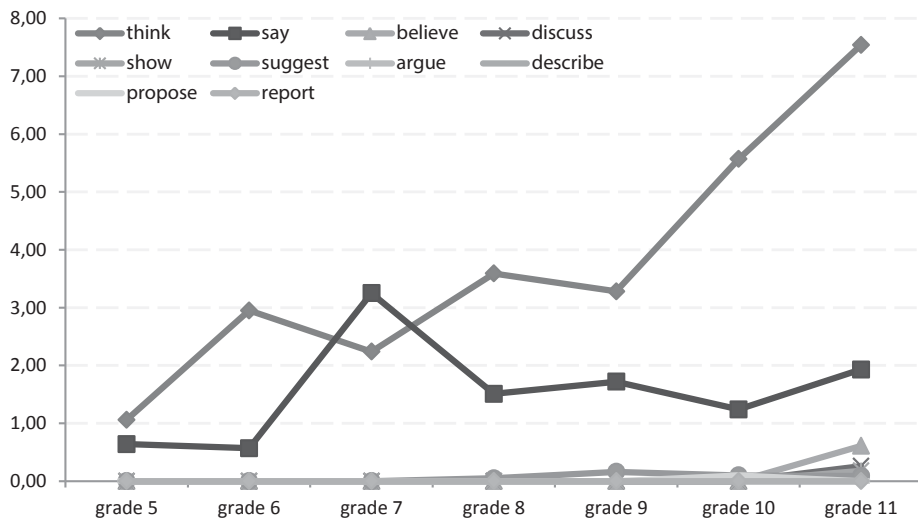
| verb | LIND- SEI | spoken. EFL | LOC- NEC | spoken. ENL | ICLE | argum. EFL | LOC- NESS | argum. ENL | CALE | acad. EFL | MIC- USP | acad. ENL |
|-----------------|--------------|----------------|-------------|----------------|------|---------------|--------------|---------------|------|--------------|-------------|--------------|
| <i>argue</i> | 0 | 0.00 | 0 | 0.00 | 21 | 0.14 | 129 | 0.61 | 39 | 0.22 | 27 | 0.43 |
| <i>believe</i> | 12 | 0.14 | 20 | 0.17 | 73 | 0.48 | 191 | 0.91 | 19 | 0.11 | 20 | 0.32 |
| <i>describe</i> | 19 | 0.22 | 3 | 0.03 | 16 | 0.10 | 19 | 0.09 | 133 | 0.76 | 41 | 0.66 |
| <i>discuss</i> | 1 | 0.01 | 7 | 0.06 | 24 | 0.16 | 49 | 0.23 | 66 | 0.38 | 59 | 0.95 |
| <i>propose</i> | 0 | 0.00 | 0 | 0.00 | 3 | 0.02 | 15 | 0.07 | 21 | 0.12 | 35 | 0.56 |
| <i>report</i> | 0 | 0.00 | 1 | 0.01 | 7 | 0.05 | 35 | 0.17 | 9 | 0.05 | 4 | 0.06 |
| <i>say</i> | 277 | 3.22 | 391 | 3.30 | 224 | 1.46 | 302 | 1.44 | 172 | 0.99 | 50 | 0.80 |
| <i>show</i> | 89 | 1.03 | 76 | 0.64 | 109 | 0.71 | 162 | 0.77 | 534 | 3.07 | 101 | 1.62 |
| <i>suggest</i> | 0 | 0.00 | 2 | 0.02 | 14 | 0.09 | 30 | 0.14 | 49 | 0.28 | 28 | 0.45 |
| <i>think</i> | 713 | 8.29 | 835 | 7.05 | 293 | 1.91 | 271 | 1.29 | 40 | 0.23 | 19 | 0.30 |

Table 4: Frequency of occurrence of ten reporting verbs per school grade in the ICCI. AUT (normalised frequencies per thousand words are given in brackets)

| verb | grade 5 | grade 6 | grade 7 | grade 8 | grade 9 | grade 10 | grade 11 |
|-----------------|----------|-----------|-----------|-----------|-----------|-----------|-----------|
| <i>argue</i> | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>believe</i> | 0 | 0 | 0 | 0 | 0 | 0 | 7 (0.61) |
| <i>describe</i> | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>discuss</i> | 0 | 0 | 0 | 0 | 0 | 0 | 3 (0.26) |
| <i>propose</i> | 0 | 0 | 0 | 0 | 0 | 1 (0.10) | 0 |
| <i>report</i> | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>say</i> | 3 (0.64) | 6 (0.57) | 74 (3.25) | 29 (1.51) | 33 (1.72) | 12 (1.24) | 22 (1.93) |
| <i>show</i> | 0 | 0 | 0 | 0 | 0 | 0 | 2 (0.18) |
| <i>suggest</i> | 0 | 0 | 0 | 1 (0.05) | 3 (0.16) | 1 (0.10) | 1 (0.09) |
| <i>think</i> | 5 (1.06) | 31 (2.95) | 57 (2.24) | 69 (3.59) | 63 (3.28) | 54 (5.57) | 86 (7.54) |

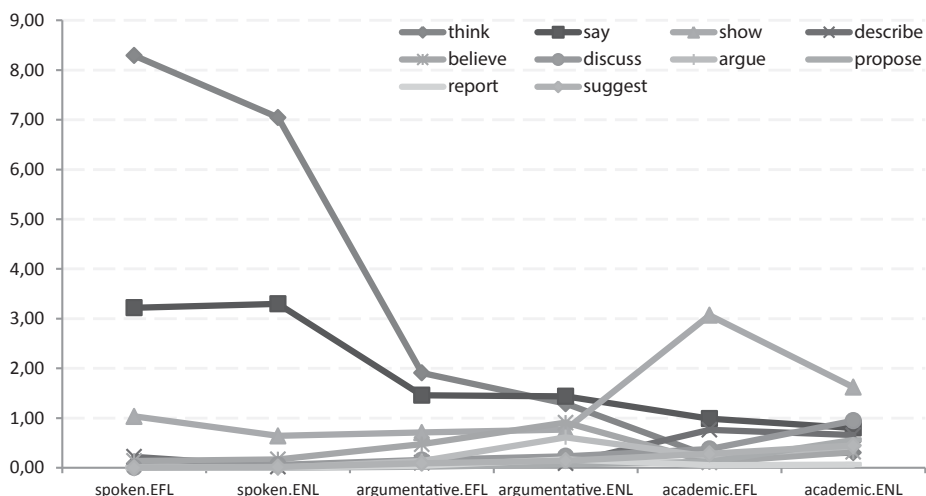
In a next step students are asked to think about visualizing the numerical data and to produce graphs by means of the spreadsheet software⁷. A standard possibility to visualize the data is shown in Figures 1 and 2 which illustrate 1) the use of *say* and *think* as core reporting verbs from early on and the slow emergence of a broader variety of such verbs from grade 5 to 11 (based on the data from the ICCLAUT), and 2) the use of the given reporting verbs across a cline from less formal (i.e. spoken) to more formal contexts (i.e. academic writing) (based on the data from the other corpora).

Figure 1: Line chart illustrating late emergence of variety of reporting verbs in grade 11 in ICCLAUT (normalised frequencies per thousand words)



⁷ In the Master's program, most students can be expected to be familiar with the basic functionality of spreadsheet software, as they will have used it in project-oriented seminars in their BA studies or when writing their BA thesis.

Figure 2: Line chart illustrating increase/decrease in frequency of individual verbs across (sub-)registers (normalised frequencies per thousand words)



Describing and explaining the findings

The students are then encouraged to analyse and interpret the findings by discussing within and between the four groups. The following are the main observations:

- In the beginning and intermediate stages of EFL learning only *say* and *think* are used as reporting verbs, most typically with first person pronouns, *say* often being used to introduce (in)direct speech; a broader variety of reporting verbs only begins to emerge in grade 11 when *believe* and *discuss* are used, albeit with low frequencies; this is partially an effect of the writing tasks set in the compilation of the ICCI as 90 % of the texts of the Austrian component examined here are descriptive/narrative (and thus, closer to speech) while argumentative texts were only written in grade 11.
- Spoken language stands out as it is dominated by the verbs *think* and *say*; only few verbs from the set of ten reporting verbs examined occur in spoken language, as many of the more formal verbs are practically absent.

- *Think*, but also *say* show a dramatic decrease in argumentative writing; here, some of the reporting verbs on the list appear, but are still used with rather low frequency counts (e.g. *describe*, *believe*, *discuss*, *argue*, *suggest*); *show* remains relatively stable across speaking and argumentative writing⁸.
- *Think* and *say* further decrease in academic writing; some other reporting verbs show an increase (e.g. *describe*, *discuss*, *show*, *suggest*, *argue*) while others decrease, most notably *believe*.
- When comparing EFL and ENL corpora, there are striking similarities in the distribution of verbs across the (sub-)registers, with only few differences in the frequency of individual verbs (e.g., *believe* and *argue* are more frequent in ENL argumentative writing; the same applies to *discuss*, *argue* and *propose* in ENL academic writing; *think* is over-represented in EFL argumentative writing, as is *show* and *say* in EFL academic writing).

The next step consists in trying to find explanations for the observed differences. Here, knowledge of some basic situational characteristics and of the communicative goals and purposes of the four (sub-)registers is helpful (if necessary, this part is facilitated by the instructor):

- spoken language (interview): interactive and ‘narrative’ purpose; speaker voices personal feelings and opinions
- descriptive writing (e.g. writing a postcard to a friend; describing oneself or one’s favourite food): interactive and ‘narrative’ purpose; speaker voices personal feelings and opinions
- argumentative writing (essay): argumentative and evaluative/persuasive purpose; writer argues for a certain position, voices a personal opinion and tries to persuade his/her (unspecified) readership
- academic writing (research paper): descriptive and informative purpose; writer describes and reports on a piece of research, his/her opinion or evaluation is usually not overt.

These situational characteristics and communicative goals and purposes should enable students to account for the predominance of *think* and *say*

8 The relatively frequent use of the verb *show* in the spoken material is due to the picture description task as part of the interviews conducted for the collection of the LINDSEI and the LOCNEC.

in spoken language and descriptive writing (as well as their sharp decrease in academic writing); the increase of *believe* and *argue* in argumentative writing; and finally, the low frequency of *think* and *believe* vs. the increasing frequency of *show* and the emergence of a broader and more formal set of verbs in academic writing. Granger and Paquot (2009) point out a major difference between argumentative and academic writing:

there is no need in argumentative writing to situate one's opinion against what has been written in the literature and typically, argumentative essays do not contain tables and graphs and are too short to include internal reference to chapters and sections (ibid.: 208).

As for the observed parallels between EFL and ENL argumentative and academic writing, it can be pointed out that these are not surprising as a) we are dealing with very advanced EFL learners and b) both groups of writers are students and thus, novice writers. However, it seems predictable that in the domain of reporting verbs, ENL writers show a higher degree of lexical variation when compared to EFL writers, and earlier corpus studies on EFL writing suggest that some of the most frequently occurring reporting verbs are in fact underrepresented in EFL writing (e.g. Granger / Rayson 1998, Neff et al. 2003, Granger / Paquot 2009). ENL writers have also been found to exhibit a more balanced use of core reporting verbs such as *say*, *state*, *show* and *argue*, whereas EFL writers more heavily rely on only one reporting verb, e.g. *say* (Neff et al. 2003: 224).

Exploring the lexico-grammatical variability of reporting verbs

The final step in the activity consists of a qualitative, exploratory look at the interaction of lexis and grammar. Not only is the choice of reporting verb closely linked to the respective text type as shown above, but reporting verbs also tend to show preferences for certain constructional variants (active voice vs. passive voice, animate vs. inanimate subject, etc.) depending on register and text type. A speaker/writer can either be overtly expressed, e.g. by using the first person pronoun *I*, or he/she can be backgrounded and even completely avoided, e.g. by using impersonal pronouns like *one* or *you*, subject placeholders (*it*, *there*), and in academic writing, a variety of inanimate nouns like *paper*, *table*, *study*, *analysis* or *findings* ~~that~~ frequently occur as subjects of active verbs (see e.g. Master 1991).

Students are asked to explore potential register differences and preferences by generating concordances for selected verbs and comparing these across the corpora. This is essentially an awareness-raising activity to help them notice the constructional variability of reporting verbs. For some verbs, students can even be asked to report differences between texts produced by learners and native speakers, and identify and explain errors, which may increase their error noticing skills (as mentioned in Section 3 above).

For the descriptive writing and the interview data, this task is relatively trivial, as the verbs *think* and *say* predominate in these (sub-)registers, and, due to the interactive and “narrative” purpose of the tasks, they are almost exclusively used in one pattern, i.e. a personal pronoun (*I, you, he/she, they*) followed by a form of *say* or *think*. For the academic register, this is more rewarding. In what follows, several patterns will be exemplified.

Argue and *propose* are examples of verbs that preferably occur in an active construction with a third-person animate subject (Callies 2013a: 378), as shown by the concordances in Figure 3. Passives are also possible, then mostly in the long passive (or *by*-passive), in which the agent is explicitly mentioned and focused on.

Figure 3: Selected concordances for *argue* and *propose* in the MICUSP

about that. Additionally, Sacks (1974) **argues** that the proposal often contains vague? (p. 12). Channell (1990) also **argues** in favor of vagueness, saying tha into the flow of conversation. Cortazzi **argues** that narratives are introduced in ks the Case feature of the verb. Lasnik **argues** that the associate in fact cannot in narrative (851). Primarily, Norrick **argues** that they have an organizational ecessarily result in three slots. Booij **proposes** that vowel length is not a pure level 1 affixes are very rare. Borowsky **proposes** that those with more than two s y principles of sonority, Dunamu (2002) **proposes** that complex onset clusters in ied the BLEU algorithm from the version **proposes** by Papineni et al., nor do they constructions, along the lines of those **proposes** by Larson (1988), as an example

Discuss is a verb that shows a much greater variability, as it is frequently used in active voice constructions with first and third person animate subjects (*I discuss..., Shaw discusses...*), but also with inanimate subjects (*This paper discusses...*) as well as in agentless passive voice constructions, some of which have formulaic character (e.g. *as discussed above/below/before/earlier*).

Figure 4: Selected concordances for discuss in the MICUSP

being overly vague. Myers (1996) also **discusses** the legitimacy of vagueness, a
 g as well, though the hedging that Shaw **discusses** is not talked of as problemati
 aker from Chemnitz, Germany. This paper **discusses** three main topics regarding t
 nted in terms of length. Both views are **discussed** below, but for the purpose of
 consonants. Word-initial consonants are **discussed** further below. Results follow
 ord order in VSO languages (p. 344). As **discussed** above, the operations that mak
 a statement, and the narrative type. As **discussed** previously, it may be that by
 vid Banks (1994a and 1994b), as will be **discussed** in more detail later, sees do
 notable exceptions and ambiguities are **discussed**. A. The GENERAL Category The
 ent, and ordering of the categories are **discussed**. Examples of each category ar
 dialects. To conclude, then, I briefly **discuss** the relations between syntax and
 t in existential constructions, where I **discuss** several arguments that have been
 ything) occupies Spec/VP? Rupp does not **discuss** this and instead focuses on
 1999). In the same article, Kegl, et al **discuss** the interactions between LSN sig

Discuss is a verb that lends itself well to a comparison of EFL and ENL data. When producing a concordance of this verb on the basis of the ICLE.GER, several instances of the ungrammatical form *discuss about* will show up that students should be able to identify when comparing them with a similar concordance from the LOCNESS (where such instances are absent). This error can be explained either by making reference to the phenomenon of cross-linguistic influence in language learning (in contrast to English, the German equivalent *diskutieren über* takes a preposition which may have been transferred) or because of analogical extension of the collocation *discussion about*, which is possible in English (see Figure 5).

Figure 5: Selected concordance lines for discuss* from the ICLE.GER

aren't far away and that it is no use to **discuss** a topic any further. Where the
 scope. Afterwards there will be an eager **discussion** about it: Whereas the "Freun
 -fashioned nappy. Two stupid mothers are **discussing** about baby's nappies; the
 so one of his defenders. This is why the **discussion** about a canon is often loade
 cluded. Lately, there have been numerous **discussions** about the establishment of
 d, indeed. After having seen the film we **discussed** about it and became all very
 uld observe you more carefully and would **discuss** about your impolite manners
 hs Augsburg's most important politicians **discuss** about this problem in the city
 soned with anabolicas. There are lots of **discussions** about today's childhood. Wh
 e support euthanasia? There are a lot of **discussions** about this problem, and it

Finally, *show* is one of several reporting verbs that predominantly takes inanimate subject, a typical feature of academic writing (see Master 1991), as exemplified in Figure 6. It seems particularly worthwhile to make students aware of such inanimate subject constructions, as these have been shown to be difficult to acquire and are thus underrepresented even in advanced L2 writing (see e.g. Callies 2013).

Figure 6: Selected concordance lines for show from the MICUSP

currence rate drops off). These results **show** that, despite the narrow criteria I
 The tables above **show** that each vowel has a slightly diff
 table below: These measurements **show** that stressed syllables have a sign
 w unresolved exceptions, this study has **shown** that, if we assume a tense/lax and
 , and study groups. The following graph **shows** the break-down of the occurrences
 e three consonants: This figure **shows** a similar trend as figure 1(the du
 verbs will be in bold. This data **shows** that while Max does see that there
 eoretic Account The rest of this paper **shows** that the data above is quite amena

5. Conclusion

In sum, I have argued in this chapter that for corpus linguistics to realize its enormous potential for data-driven and research-based language learning and teaching, both pre-service teacher trainees and experienced, qualified teachers must be provided with corpus literacy. As an example of how corpus linguistics can be integrated into curricula for foreign language teachers, I have described an advanced, student-centered DDL-activity that makes use of various types of native-language and learner corpora to raise teacher students' awareness of register differences and aspects of the lexicogrammatical variability of reporting verbs in English. The activity draws on several subcomponents of corpus literacy, such as searching corpora by means of concordancers, as well as processing, analysing and interpreting corpus data. Learner corpora can be integrated to increase future teachers' error awareness and noticing skills.

Students of foreign languages are increasingly confronted with linguistic corpora in their studies. But for them to acquire corpus literacy to a degree that they will consider corpora as part of the toolbox for language teachers, we need to develop and implement specialized linguistic modules for language teacher students that train them to use and exploit corpora

to actually teach language. To be able to do this, corpus linguists need to collaborate with experts in foreign language didactics and pedagogy, and with teaching practitioners as well. For in-service teachers, workshops or online teacher-training courses on applied corpus linguistics (e.g. Leńko-Szymańska 2015) seem promising options.

As a spin-off from the DDL-activity outlined in this chapter, the development of an interactive online writing tool could be envisaged. This tool would combine information about the characteristics and use of reporting verbs with real examples drawn from native-language and learner corpora. Users should be able to access information in two ways:

- form-to-function, i.e. looking up individual verbs in an alphabetical index to see how they are used (indicating typical object- and subject-noun collocates of a verb as well as its preferred constructional patterns)
- function-to-form, i.e. accessing a list of functions (e.g. expressing a positive/neutral/negative attitude towards the truth of a claim) to find verbs typically used to express this function.

Such a tool would expand on and complement existing web resources such as the “Academic Phrasebank” (2015) or Bloch’s (2015) reporting verbs tutorial.

References

- Academic Phrasebank. Retrieved 30.7.2015, from <http://www.phrasebank.manchester.ac.uk/>.
- Anthony, L., 2015: *AntConc* (Version 3.4.4). (Computer Software) Waseda University: Tokyo. Retrieved 30.7.2015, from <http://www.laurenceanthony.net/>.
- Biber, Douglas. et al., 1999: *Longman Grammar of Spoken and Written English*. Longman: Harlow.
- Biber, Douglas / Conrad, Susan / Leech, Geoffrey, 2002: *Longman Student Grammar of Spoken and Written English*. Longman: Harlow.
- Bloch, Joel 2010: “A concordance-based study of the use of reporting verbs as rhetorical devices in academic papers”. *Journal of Writing Research* 2(2), 219–244.

- Bloch, Joel, 2015: *Reporting Verb Recourse Site*. Retrieved 30.7.2015, from <http://eslcomposition.osu.edu/>.
- Breyer, Yvone, 2009: "Learning and teaching with corpora: reflections by student teachers". *Computer Assisted Language Learning* 22(2), 153–172.
- Callies, Marcus, 2013: "Agentivity as a determinant of lexico-grammatical variation in L2 academic writing". *International Journal of Corpus Linguistics* 18(3), 357–390.
- Callies, Marcus / Zaytseva, Ekaterina, 2013: "The *Corpus of Academic Learner English* (CALE) – A new resource for the assessment of writing proficiency in the academic register". *Dutch Journal of Applied Linguistics* 2(1), 126–132.
- Campoy-Cubillo, Mari Carmen / Bellés-Fortuño, Begona / Gea-Valor, Maria Lluïsa (eds.), 2010: *Corpus-Based Approaches to English Language Teaching*. Continuum: London/New York.
- Dalton-Puffer, Christiane, 2014: *Corpus Linguistics in language teacher education*. (Plenary lecture given at the 14th Klagenfurt Conference on Corpus Based Applied Linguistics (CALK 14), 25–27 September 2014). University of Klagenfurt, Austria.
- Deutscher Anglistenverband und Deutsche Gesellschaft für Amerikastudien, 2009: *Inhaltliche Anforderungen für Fachwissenschaft und Fachdidaktik in der Lehrerinnen- und Lehrerbildung, Studienfach Englisch*. (Positionspapier). Retrieved 29.7.2015, from http://www.anglistenverband.de/wp-content/uploads/LA-Englisch_Positionspapier.pdf.
- Farr, Fiona, 2008: "Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users". *Language Awareness* 17(1), 25–43.
- Flowerdew, Lynne, 2012: *Corpora and Language Education*. Palgrave: Basingstoke.
- Frankenberg, Ana / Flowerdew, Lynne / Aston, Guy (eds.), 2011: *New Trends in Corpora and Language Learning*. London: Continuum.
- Gilquin, Gaëtanelle/ De Cock, Sylvie / Granger, Sylviane, 2010: *The Louvain International Database of Spoken English Interlanguage*. Louvain UP: Louvain-la-Neuve.

- Granger, Sylviane / Rayson, Paul, 1998: "Automatic profiling of learner texts". In: Granger, S. (ed.): *Learner English on Computer*. Longman: London/New York, 119–131.
- Granger, Sylviane/ Paquot, Magali, 2009: "Lexical verbs in academic discourse: A corpus-driven study of learner use". In: Charles, Maggie / Pecorari, Diane / Hunston, Susan (eds.): *Academic Writing. At the Interface of Corpus and Discourse*. Continuum: London/New York, 193–214.
- Granger, Sylviane, 2009: "The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation". In: Aijmer, Karin (ed.): *Corpora and Language Teaching*. Benjamins: Amsterdam, 13–33.
- Granger, Sylviane et al. (eds.), 2009: *International Corpus of Learner English V2*. Louvain UP: Louvain-la-Neuve.
- Heather, Julian / Helt, Marie, 2012: "Evaluating corpus literacy training for pre-service language teachers: Six case studies". *Journal of Technology and Teacher Education* 20(4), 415–440.
- Hyland, Ken, 1999: "Academic attribution. Citation and the construction of disciplinary knowledge". *Applied Linguistics* 20(3), 341–367.
- James, Thomas / Boulton, Alex (eds.), 2012: *Input, Process and Product: Developments in Teaching and Language Corpora*. Masaryk UP: Brno.
- Kübler, Natalie (ed.), 2011: *Corpora, Language, Teaching, and Resources. From theory to practice*. Peter Lang: Bern.
- Leńko-Szymańska, Agnieszka, 2014a: "Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education". *ReCALL* 26(2), 260–278.
- Leńko-Szymańska, Agnieszka, 2014b: "A teacher-training course on the use of corpora in language education: Perspectives of the students". In: Turla, Anna / Mikołajewska, Beata / Stanulewicz, Danuta (eds.): *Insights into Technology-enhanced Language Pedagogy*. Peter Lang: Frankfurt am Main, 129–144.
- Leńko-Szymańska, Agnieszka, 2015: *Corpora in Language Teaching and Learning*. Retrieved 29.7.2015, from <http://corpora.blog.ils.uw.edu.pl/>.
- Leńko-Szymańska, Agnieszka / Boulton, A. (eds.), 2015: *Multiple Affordances of Language Corpora for Data-driven Learning*. Benjamins: Amsterdam.

- Master, Peter, 1991: "Active verbs with inanimate subjects in scientific prose". *English for Specific Purposes* 10(1), 15–33.
- Mukherjee, Joybrato, 2004: "Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany". In: Connor, Ulla / Upton, Thomas A. (eds.): *Applied Corpus Linguistics. A Multi-dimensional Perspective*. Rodopi: Amsterdam, 239–250.
- Mukherjee, Joybrato, 2006: "Corpus linguistics and language pedagogy: the state of the art – and beyond". In: Braun, Sabine / Kohn, Kurt / Mukherjee, Joybrato (eds.): *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Peter Lang: Frankfurt am Main, 5–24.
- Neff, JoAnne, Dafouz et al, 2003: "Contrasting learner corpora: The use of modal and reporting verbs in the expression of writer stance". In: Granger, Sylviane / Petch-Tyson, Stephanie (eds.): *Extending the Scope of Corpus-Based Research. New Applications, New Challenges*. Rodopi: Amsterdam, 211–230.
- Nesselhauf, Nadja, 2004: "Learner corpora and their potential for language teaching". In: Sinclair, John (ed.): *How to Use Corpora in Language Teaching*. Benjamins: Amsterdam, 125–152.
- O’Keeffe, Anne / McCarthy, Michael (eds.), 2010: *The Routledge Handbook of Corpus Linguistics*. Routledge: New York.
- Reppen, Randi, 2010: *Using Corpora in the Language Classroom*. Cambridge UP: Cambridge.
- Römer, Ute, 2009: "Corpus research and practice: What help do teachers need and what can we offer?" In: Aijmer, Karin (ed.): *Corpora and Language Teaching*. Benjamins: Amsterdam, 83–98.
- Römer, Ute, 2011: "Corpus research applications in second language teaching". *Annual Review of Applied Linguistics* 31, 205–225.
- Römer, Ute / O’Donnell, Matthew Brook, 2011: "From student hard drive to web corpus (part 1): The design, compilation and genre classification of the Michigan Corpus of Upper-level Student Papers (MICUSP)". *Corpora* 6(2), 159–177.
- Scott, Mike, 2008: *WordSmith Tools version 5*. (Computer Software) Lexical Analysis Software: Liverpool.
- Shaw, Philip, 1992: "Reasons for the correlation of voice, tense, and sentence function in reporting verbs". *Applied Linguistics* 13(3), 302–319.

- Thomas, Sarah / Hawes, Thomas P., 1994: “Reporting verbs in medical journal articles”. *English for Specific Purposes* 13(2), 129–148.
- Thompson, Geoff / Yiyun, Ye, 1991: “Evaluation in the reporting verbs used in academic papers”. *Applied Linguistics* 12(4), 365–382.
- Tribble, Chris, 2015: “Teaching and language corpora. Perspectives from a personal journey”. In: Leńko-Szymańska, Agnieszka / Boulton, Alex (eds.): *Multiple Affordances of Language Corpora for Data-driven Learning*. Benjamins: Amsterdam, 37–62.
- Tono, Yukio / Díez-Bedmar, Maria Belen, 2014: “Focus on learner writing at the beginning and intermediate stages. The ICCI corpus”. *International Journal of Corpus Linguistics* 19(2), 163–177.
- University of Bremen, 2015: “Module description for the linguistics module LING in the Master of Education / English program”. Retrieved 29.7.2015, from <http://www.fb10.uni-bremen.de/anglistik/ba2/module/LING.pdf>.