

# ***The Corpus of Academic Learner English (CALE)* – A new resource for the study and assessment of advanced language proficiency**

Marcus Callies, Ekaterina Zaytseva  
University of Bremen

## **Abstract**

This paper introduces the *Corpus of Academic Learner English (CALE)*, a Language for Specific Purposes learner corpus that is currently being compiled for the quantitative and qualitative study of advanced learners' written academic English. CALE is designed to comprise seven academic genres produced by learners of English as a foreign language in a university setting and thus contains discipline- and genre-specific texts. The corpus will serve as an empirical basis to produce detailed case studies that examine linguistic determinants of lexico-grammatical variation, *i.e.* semantic, structural, discourse-motivated and processing-related factors that influence constituent order and the choice of structural variants, but also those that are potentially more specific to the acquisition of L2 academic writing such as task setting, genre and writing proficiency. Another major goal is to develop a set of linguistic criteria for the assessment of advanced proficiency conceived of as "sophisticated language use in context".

**Keywords:** learner English, academic writing, lexico-grammatical variation, advanced proficiency.

## **1. Introduction**

Recently, Second Language Acquisition (SLA) research has seen an increasing interest in advanced stages of acquisition and questions of near-native competence. Learner Corpus Research (LCR) has contributed to a much clearer picture of advanced interlanguages, providing evidence that learners of various native language (L1) backgrounds have similar problems and face similar challenges on their way to near-native proficiency. Despite the growing interest in advanced proficiency, the fields of SLA and LCR are still struggling with (1) a definition and clarification of the concept of "advancedness", (2) an in-depth description of advanced proficiency, especially as to the acquisition of optional and highly L2-specific phenomena which are often located at the interfaces of linguistic subfields, and (3) the operationalization of such a description in terms of criteria for the assessment of advancedness. In this paper, we introduce the *Corpus of Academic Learner English (CALE)*<sup>1</sup>, a Language for Specific Purposes learner corpus that is currently being compiled for the quantitative and qualitative study of patterns and determinants of lexico-grammatical variation in advanced learners' written academic English.

---

<sup>1</sup> More information about the corpus project can be accessed at <http://www.advanced-learner-varieties.info/> (last accessed on 12 March, 2013).

## 2. Corpus design, composition and annotation

Many existing and widely-used learner corpora, such as the *International Corpus of Learner English* (ICLE, Granger *et al.* 2009) include learner writing of a general argumentative, creative or literary nature, and thus not academic writing in a narrow sense. Academic writing in our context is conceived of as

any writing that fulfills a purpose of education in a college or university (...); writing in response to an academic assignment, or professional writing that trained 'academics' – teachers and researchers – do for publications read and conferences attended by other academics. (Thaiss & Zawacki 2006: 4)

Biber & Conrad (2009) consider academic prose

a very general *register*, characterized as written language that has been carefully produced and edited, addressed to a large number of readers who are separated in time and space from the author, and with the primary communicative purpose of presenting information about some topic. (Biber & Conrad 2009: 32; our emphasis)

Several patterns of variation that predominantly occur in academic prose, or are subject to the characteristic features of this register, are not represented at all or not frequently enough in general learner corpora. CALE is designed to comprise academic texts produced by learners of English as a foreign language (EFL) in a university setting. CALE may therefore be considered a Language for Specific Purposes learner corpus, containing discipline- and genre-specific texts (Granger & Paquot 2012). Similar corpora that contain native speaker (NS) writing and may thus serve as control corpora for CALE are the *Michigan Corpus of Upper-Level Student Papers* (MICUSP, Römer & Brook O'Donnell 2011) and the corpus of *British Academic Written English* (BAWE, Alsop & Nesi 2009).

CALE comprises seven academic text types ("genres") produced as assignments by EFL learners in university content courses (linguistics), see Figure 1. Currently, we are mostly collecting texts and bio data from German students, but are planning to include data from EFL learners of other L1 backgrounds to be able to make cross-linguistic and typological comparisons as to potential L1 influence. Students who are submitting texts for the corpus can be considered 'advanced' in terms of external criteria, *i.e.* the time spent learning English at school and university. There are several problems connected with the use of such external criteria to globally assess proficiency. The assessment of proficiency in general is a notoriously difficult (and also frequently neglected and underestimated) challenge in SLA and LCR. It is not possible to go into detail here, but see Callies & Zaytseva (2013) for further discussion as to the operationalization and assessment of advancedness in the CALE project.

The text classification we have developed for CALE is comparable with the NS control corpora mentioned above, but we have created clear(er) textual profiles, adopting the situational characteristics and linguistic features identified for academic prose by Biber & Conrad (2009). The communicative purpose or goal of the text serves as the main classifying principle, which helps to set apart the seven genres in terms of (a) the general purpose of the text, (b) its specific purpose(s), (c) the skills the author demonstrates, and (d) the author's stance. In addition, we list the major features

of each text type as to (a) structural features, (b) length, and (c) functional features. Table 1 illustrates the profile for the academic genre "abstract".

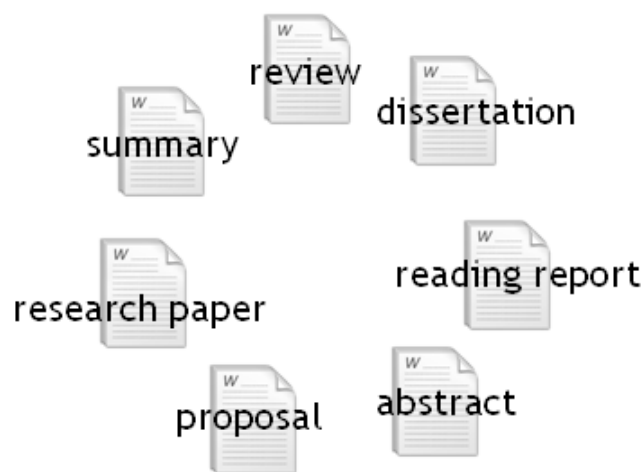


Figure 1. Academic text types in CALE

Communicative goal/purpose	Features
<p><i>a. general purpose</i> informational – to inform</p> <p><i>b. specific purpose(s)</i> captures the essence of published research (why, how, what: research focus, methodology, results/findings, conclusion &amp; recommendations); should help the reader to quickly ascertain purpose, content and usefulness of publication</p> <p><i>c. skills</i> author demonstrates ability to extract and provide essential information in an exhaustive and compelling way</p> <p><i>d. stance</i> author’s opinion/evaluation is absent</p>	<p><i>a. structural</i> not structured into sections; appears at beginning of text it comes with; may also occur as stand-alone entity instead of full paper</p> <p><i>b. length</i> rather short (approx. 100-250 words), rarely exceeding 500 words</p> <p><i>c. functional</i> self-contained piece of writing, can be understood independently from accompanying publication</p>

Table 1. Profile for the text type "abstract" in CALE

Students submit their texts in electronic form (typically in .doc, .docx or .pdf file format). Thus, some manual pre-processing of these incoming files is necessary. Extensive "non-linguistic" information (such as table of contents, list of references, tables and figures, *etc.*) is deleted and substituted by placeholder tags around their headings or captions. The body of the text is then annotated for meta-textual, *i.e.* underlying structural features (section titles, paragraphs, quotations, examples, *etc.*)

with the help of annotation tools. The texts are also annotated (in a file header) for metadata, *i.e.* learner variables such as L1, age, gender, *etc.* which are collected through a written questionnaire. The file header also includes metadata that pertain to each individual text such as genre, type of course and discipline the text was written in, the setting in which the text was produced, *etc.* This information is also collected with the help of a questionnaire that accompanies each text submitted to the corpus.

Based on a function-to-form approach to the analysis of learner language, (parts of) the corpus will be annotated for linguistic features on an extra layer, *e.g.* for rhetorical functions like contrast and the lexico-grammatical means to express it (Zaytseva 2011). The advantages of a function-driven type of annotation of learner corpora are clearly at hand. It establishes a useful link between the occurrence of particular linguistic structures and the functions that they serve in discourse, accounting for unexpected, covert structures that are bound to occur in learner language but will often be overlooked in form-driven searches. In this approach, learners are not merely considered non-native speakers who demonstrate deficient L2 knowledge, but active language users who may (creatively) use non-canonical, strategic means to express communicative functions (*cf.* Hirschmann *et al.* 2007; Lüdeling 2008). A good example of such an unexpected, strategic behaviour is German EFL learners' use of questions to highlight information, as pointed out by Callies (2009: 138). Moreover, while form-based annotation and searching often singles out a subset of the most typically occurring structures, a function-driven annotation makes possible the identification and documentation of a near-complete inventory of lexico-grammatical and syntactic means used to express various communicative functions in written discourse. Crucially, this approach leads to corpus searches that retrieve a stock of constructions in one query, while in a form-oriented approach, many more, partially very complex queries are needed that are additionally subject to extensive manual post-processing due to a potentially low precision rate.

### 3. Research program

In the following sections, we outline our research program. We adopt a variationist perspective on SLA, combining a learner corpus approach with research on near-native competence.

#### 3.1. The study of variation in SLA research

Interlanguages (ILs) as varieties in their own right are characterized by variability and instability even more than native languages. Research on IL-variation since the late 1970s has typically focused on beginning and intermediate learners and the (un-)successful learning of actually invariant linguistic forms in pronunciation and morphosyntax, *i.e.* the occurrence of alternations between target-like and non-target-like equivalent forms. Such studies revealed developmental patterns, interpreted as indicators of learners' stages of acquisition, and produced evidence that IL-variation co-varies with linguistic, social/situational and psycholinguistic context, and is also subject to a variety of other factors like individual learner characteristics and biographical variables (*e.g.* form and length of exposure to the L2). Since the early 2000s there has been an increasing interest in issues of sociolinguistic and sociopragmatic variation in advanced L2 learners, frequently referred to as

sociolinguistic competence, *e.g.* learners' use of dialectal forms or pragmatic markers (mostly in L2 French, see *e.g.* Mougeon & Dewaele 2004; Regan *et al.* 2009). This has marked both a shift from the study of beginning and intermediate to advanced learners, and a shift from the study of norm-violations to the investigation of differential knowledge as evidence of conscious awareness of (socio-)linguistic variation.

### 3.2. Advanced Learner Varieties (ALVs)

There is evidence that advanced learners of various language backgrounds have similar problems and face similar challenges on their way to near-native proficiency. In view of these assumed similarities, some of which will be discussed in the following, we conceive of the interlanguage of these learners as Advanced Learner Varieties (ALVs). In a recent overview of the field, Granger (2008: 269) defines advanced (written) interlanguage as "the result of a highly complex interplay of factors: developmental, teaching-induced and transfer-related, some shared by several learner populations, others more specific". According to her, typical features of ALVs are overuse of high frequency vocabulary and a limited number of prefabs, a much higher degree of personal involvement, as well as stylistic deficiencies, "often characterized by an overly spoken style or a somewhat puzzling mixture of formal and informal markers" (Granger 2008: 269). Moreover, advanced learners typically struggle with the acquisition of optional and highly L2-specific linguistic phenomena, often located at interfaces of linguistic subfields (*e.g.* syntax-semantics, syntax-pragmatics, see *e.g.* DeKeyser 2005: 7ff.). As to academic writing, many of their observed difficulties are caused by a lack of understanding of the conventions of academic writing, or a lack of practice, but are not necessarily a result of interference from L1 academic conventions (McCrostie 2008: 112).

### 3.3. Patterns and determinants of variation in L2 (academic) writing

Our research program involves the study of L2 learners' acquisition of the influence of several factors on constituent order and the choice of constructional variants (*e.g.* verb-particle placement, focus constructions). One reason for this is that such variation is often located at the interfaces of linguistic subsystems, an area where advanced learners still face difficulties. Moreover, grammatical variation in L2 has not been well researched to date and is only beginning to attract researchers' attention (Gries & Wulff 2005, 2009, 2011; Callies 2008, 2009; Callies & Szczesniak 2008; Wulff & Gries 2011). There are a number of semantic, structural, discourse-motivated and processing-related determinants that influence lexico-grammatical variation whose interplay and influence on speakers' and writers' constructional choices has been widely studied in corpus-based research on L1 English. Generally speaking, in L2 English these determinants play together with several IL-specific ones such as L1 and proficiency level, and in writing in particular, some further task-specific factors like imagined audience (the people to whom the text is addressed), task setting, register and genre add to this complex interplay of factors, see Figure 2.

<b>Advanced Learners</b>	<b>Native speakers + Advanced Learners</b>	
<b>IL-specific</b>	<b>context/task-specific</b>	<b>linguistic</b>
<ul style="list-style-type: none"> <li>- mother tongue (L1)</li> <li>- target language (L2)</li> <li>- (writing) proficiency</li> <li>- teaching/instruction</li> <li>- ...</li> </ul>	<ul style="list-style-type: none"> <li>- imagined audience</li> <li>- task-type</li> <li>- task-setting</li> <li>- genre</li> <li>- ...</li> </ul>	<ul style="list-style-type: none"> <li>- structure</li> <li>- processing</li> <li>- semantics</li> <li>- discourse</li> <li>- ...</li> </ul>

Figure 2. Determinants of variation in L1 and L2 writing

In the following, two examples and a case study reported on in Section 3.4 will serve to illustrate how such factors influence lexico-grammatical variation. To begin with, examples (1) and (2) from CALE show the variable use of transitive phrasal verbs in English: they allow the particle to occur either before or after the object NP.

1. After having checked the OED for already established formations, I **sorted** eight words of the fiction section **out** and another four from the section non-fiction which were attested before the 20th century. However, I kept those words which were attested by the OED in the 20th century and only **sorted out** those that were coined before. (CALE)
2. As magazines show a remarkably smaller amount of *-ee* words than newspapers, we have to **think** the statement made in analysis (1) **over** that one could assume to find more *-ee* words in magazines because of their more whimsical tone. (CALE)

Thus, the word order can alternate between a split ordering in which the particle is separated from the verb and follows the object NP (*I sorted eight words of the fiction section out*), and a joined ordering in which both verb and particle precede the object NP (*I [...] only sorted out those that were coined before*). Several determinants have been shown to influence this variation in L1 English, many of which can be related to processing effort (see *e.g.* Gries 2002 for a unified account), and it is well-known that one of the most important factors that determines the preferred positioning of the particle is the weight (in the sense of length and/or complexity) of the object NP: when the object is heavy, the joined ordering is preferred. All three instances of particle placement in (1) and (2) are cases in which the object NP can be said to be heavy (and complex) in that it is postmodified either by a prepositional phrase, or a full or truncated relative clause, thus preferring the joined ordering. Interestingly, however, this preferred order is chosen in only one of these instances, suggesting that the way in which weight/complexity influences constituent order in advanced learner varieties does merit closer attention.

Weight effects are also at play in example (3a). Here, a learner uses a premodified NP which is thematically packed and as a consequence, unnaturally heavy when compared to the improved re-ordering in (3b) (see also Lorenz 1999 for similar findings in the context of adjective intensification).

- 3.a These differences that <quote removed> constitute **subtle, but through corpus research nevertheless traceable deviations** from an implied, abstract native-speaker norm. (CALE)
- 3.b These differences that <quote removed> constitute subtle deviations from an implied, abstract native-speaker norm, but are nevertheless traceable through corpus research.

Turning to the interplay of IL- and task-specific factors, in corpus-based studies differences between texts produced by L1 and L2 writers are often attributed to the influence of the learners' L1. However, such differences may in fact turn out to result from differences in task-setting (*e.g.* prompt, timing, access to reference works, see Ädel 2008), and possibly task-instruction and imagined audience (see Ädel 2006: 201ff. for a discussion of corpus comparability). Similarly, research findings as to learners' use of features that are more typical of speech than of academic prose have been interpreted as unawareness of register differences, but there is some evidence that the occurrence of such forms may also be caused by the influence of factors like the development of writing proficiency over time (novice writers *vs.* experts, see Gilquin & Paquot 2008), task-setting and -instruction (Ädel 2008), imagined audience, or text type and (sub-)register (*e.g.* text type-specific conventions, see Wulff & Römer 2009, or academic *vs.* argumentative writing, see Zaytseva 2011).

#### 3.4. Case study: agentivity as a determinant of lexico-syntactic variation in L2 academic writing

In this section, we provide another example of how linguistic variation plays out in L2 academic writing. In a CALE pilot study of the (non-) representation of authorship in research papers written by advanced German EFL learners, Callies (2013) examined agentivity as a determinant of lexico-syntactic variation in academic prose. He hypothesized that even advanced students are insecure about the representation of authorship due to a mixture of several reasons: conflicting advice by teachers, textbooks and style guides, the diverse conventions of different academic disciplines, students' relative unfamiliarity with academic text types and lack of linguistic resources to report events and findings without mentioning an agent. Interestingly, the study found both an overrepresentation of the first person pronouns *I* and *we*, as well as an overrepresentation of the highly impersonal subject-placeholders *it* and *there* (often used in the passive voice) in the learner texts as default strategies to suppress the agent, see examples (4) and (5).

4. **There are two things to be discussed** in this section. (CALE)
5. **It can be observed that** the other neologisms which only consist of one syllable have an insertion of a linking vowel. (CALE)

While this finding seems to be contradictory, it can be explained by a third major finding, namely the significant underrepresentation of inanimate subjects which are, according to Biber & Conrad (2009: 162), preferred reporting strategies in L1 academic English, as exemplified in (6).

6. **This seminar paper aims to investigate** subject verb concord with collective nouns. (CALE)

Callies concluded that L2 writers have a narrower inventory of linguistic resources to report events and findings without an overt agent, and their insecurity and unfamiliarity with academic texts adds to the observed imbalanced clustering of first person pronouns, dummy-subjects and passives. The findings of this study also suggest that previous studies that frequently explain observed overrepresentations of informal, speech-like features by pointing to learners' higher degree of subjectivity and personal involvement (Granger 2008) or unawareness of register differences (Gilquin & Paquot 2008), may need to be supplemented by taking into account a more complex interplay of factors that also includes the limited choice of alternative strategies available to L2 writers.

#### **4. Implications for language teaching, testing and assessment**

The project we have outlined in this paper has some major implications for EFL teaching and assessment. The research findings will be used to provide recommendations for EFL teachers and learners by developing materials for teaching units in practical language courses on academic writing and English for Academic Purposes. In the long run, we plan to create a web-based reference tool that will help students look up typical collocations and recurring phrases used to express rhetorical moves/functions in academic writing (*e.g.* giving examples, expressing contrast, drawing conclusions, *etc.*). This application will be geared towards students' needs and can be used as a self-study reference tool at all stages of writing an academic text. Users will be able to access information in two ways: 1) form-to-function, *i.e.* looking up words and phrases in an alphabetical index to see how they can express rhetorical functions, and 2) function-to-form, *i.e.* accessing a list of rhetorical functions to find words and phrases that are typically used to encode them. Most importantly, the tool will present in a comparative manner structures that emerged as problematic in advanced learners' writing, for example untypical lexical co-occurrence patterns and over- or underrepresented words and phrases, side by side with those structures that typically occur in expert academic writing. This will include information on the immediate and wider context of use of single items and multi-word-units. While the outcome is thus particularly relevant for future teachers of English, it may also be useful for students and academics in other disciplines who have to write and publish in English. Unlike in the Anglo-American education system, German secondary schools and universities do not usually provide courses in academic writing in the students' mother tongue, so that first-year students have basically no training in academic writing at all.

It has been mentioned earlier that the operationalization of a quantitatively and qualitatively well-founded description of advanced proficiency in terms of criteria for the assessment of advancedness is still lacking. Thus, a major aim of the project is to develop a set of linguistic descriptors for the corpus-based assessment of advanced proficiency. Learner corpora are increasingly used for language testing and assessment (Taylor & Barker 2008, Barker 2010) and have the potential to increase transparency, consistency and comparability in the assessment of L2 proficiency. However, learners' proficiency level has been a fuzzy variable in LCR so far in that it has often been assessed globally by means of external criteria, *i.e.* institutional status ("university undergraduates in English, usually in their third or fourth year", Granger *et al.* 2009: 10). As a consequence, in some learner corpora the proficiency level varies between



higher intermediate and advanced, thus giving rise to the "across-the-board" problem when research findings are reported: existing individual learner differences tend to be disregarded and reported average frequency counts may be skewed.

The observation that the descriptors and can-do-statements of the *Common European Framework of Reference* (CEFR; Council of Europe 2001) often appear too global and general to be of practical value for language assessment, and are underspecified for describing advanced learners' competence in academic writing, has led to an increasing awareness of the need to develop linguistic descriptors (Neff van Aertselaer & Bunce 2011) or "criterial features" (Hawkins & Buttery 2010). Ortega & Byrnes (2008) discuss four ways in which advancedness has commonly been operationalised, ultimately favouring what they call "sophisticated language use in context", a construct that includes *e.g.* the choice among registers, repertoires and voice. This concept will serve as a basis for the development of linguistic descriptors that are characteristic of academic prose (see Callies & Zaytseva 2013), *e.g.* the use of syntactic structures like inanimate subjects, phrases to express rhetorical functions (*e.g. by contrast, to conclude, in fact*), reporting verbs (*discuss, claim, suggest, argue, propose, etc.*), and lexical co-occurrence patterns (*e.g. conduct, carry out and undertake* as typical verbal collocates of *experiment, analysis* and *research*).

## References

- Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. Amsterdam: Benjamins.
- Ädel, A. (2008). Involvement features in writing: Do time and interaction trump register awareness? In G. Gilquin, S. Papp & M.B. Díez-Bedmar (eds) *Linking up Contrastive and Learner Corpus Research*. Amsterdam: Rodopi, 35-53.
- Alsop, S. & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora* 4(1), 71-83.
- Barker, F. (2010). How can corpora be used in language testing? In A. O'Keeffe & M. McCarthy (eds) *The Routledge Handbook of Corpus Linguistics*. New York: Routledge, 633-645.
- Biber, D. & Conrad, S. (2009). *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Callies, M. (2008). Easy to understand but difficult to use? Raising constructions and information packaging in the advanced learner variety. In G. Gilquin, S. Papp & M.B. Díez-Bedmar (eds) *Linking up Contrastive and Learner Corpus Research*. Amsterdam: Rodopi, 201-226.
- Callies, M. (2009). *Information Highlighting in Advanced Learner English*. Amsterdam: Benjamins.
- Callies, M. (2013). Agentivity as a determinant of lexico-syntactic variation in L2 academic writing. *International Journal of Corpus Linguistics* 18(3).
- Callies, M. & Szczesniak, K. (2008). Argument realization, information status and syntactic weight - A learner-corpus study of the dative alternation. In M. Walter & P.

Grommes (eds) *Fortgeschrittene Lernervarietäten. Korpuslinguistik und Zweitspracherwerbsforschung*. Tübingen: Niemeyer, 165-187.

Callies, M. & Zaytseva, E. (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).

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press. [http://www.coe.int/t/dg4/linguistic/Source/Framework\\_en.pdf](http://www.coe.int/t/dg4/linguistic/Source/Framework_en.pdf) (last accessed on 12 March, 2013).

DeKeyser, R. (2005). What makes learning second language grammar difficult? A review of issues. *Language Learning* 55(1), 1-25.

Gilquin, G. & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction* 1(1), 41-61.

Granger, S. (2008). Learner corpora. In A. Lüdeling & M. Kytö (eds) *Corpus Linguistics. An International Handbook, Vol. 1*. Berlin & New York: Mouton de Gruyter, 259-275.

Granger, S., Dagneaux, E., Meunier, F. & Paquot, M. (2009). *The International Corpus of Learner English. Version 2. Handbook and CD-ROM*. Louvain-la-Neuve: Presses universitaires de Louvain.

Granger, S. & Paquot, M. (2012). Language for specific purposes learner corpora. In T.A. Upton & U. Connor (eds) *Language for Specific Purposes. The Encyclopedia of Applied Linguistics*. New York: Blackwell, 3142-3146.

Gries, S.Th. (2002). The influence of processing on grammatical variation: Particle placement in English. In N. Dehé, R. Jackendoff, A. McIntyre & S. Urban (eds) *Verb-Particle Explorations*. Berlin: Mouton de Gruyter, 269-288.

Gries, S.Th. & Wulff, S. (2005). Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics* 3, 182-200.

Gries, S.Th. & Wulff, S. (2009). Psycholinguistic and corpus-linguistic evidence for L2 constructions. *Annual Review of Cognitive Linguistics* 7, 164-187.

Gries, S.Th. & Wulff, S. (2011). Constituent-order alternation phenomena in L2: Two multifactorial and processing-based case studies. Paper presented at *Learner Corpus Research 2011*, Université catholique de Louvain, Belgium, 15-17 September, 2011.

Hawkins, J. & Buttery, P. (2010). Criterial features in learner corpora: Theory and illustrations. *English Profile Journal* 1(1), 1-23.

Hirschmann, H., Doolittle, S. & Lüdeling, A. (2007). Syntactic annotation of non-canonical linguistic structures. In M. Davies, P. Rayson, S. Hunston & P. Danielsson (eds) *Proceedings of Corpus Linguistics 2007*, Birmingham.

- Lorenz, G. (1999). *Adjective Intensification – Learners Versus Native Speakers. A Corpus Study of Argumentative Writing*. Amsterdam: Rodopi.
- Lüdeling, A. (2008). Mehrdeutigkeiten und Kategorisierung: Probleme bei der Annotation von Lernerkorpora. In M. Walter & P. Grommes (eds) *Fortgeschrittene Lernervarietäten. Korpuslinguistik und Zweitspracherwerbsforschung*. Tübingen: Niemeyer, 119-140.
- McCrostie, J. (2008). Writer visibility in EFL learner academic writing: A corpus-based study. *ICAME Journal* 32, 97-114.
- Mougeon, R. & Dewaele, J.-M. (eds). (2004). Patterns of variation in the interlanguage of advanced second language learners. Special issue of *International Review of Applied Linguistics in Language Teaching (IRAL)* 42(4), 295-402.
- Neff van Aertselaer, J. & Bunce, C. (2011). The use of small corpora for tracing the development of academic literacies. In F. Meunier, S. De Cock, G. Gilquin & M. Paquot (eds) *A Taste for Corpora*. Amsterdam: Benjamins, 63-84.
- Ortega, L. & Byrnes, H. (2008). The longitudinal study of advanced L2 capacities: An introduction. In L. Ortega & H. Byrnes (eds) *The Longitudinal Study of Advanced L2 Capacities*. New York: Routledge/Taylor & Francis, 3-20.
- Regan, V., Howard, M. & Lemée, I. (2009). *The Acquisition of Sociolinguistic Competence in a Study Abroad Context*. Clevedon: Multilingual Matters.
- Römer, U. & Brook O'Donnell, M. (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.
- Taylor, L. & Barker, F. (2008) Using corpora for language assessment. In E. Shohamy & N.H. Hornberger (eds) *Encyclopedia of Language and Education, 2nd Edition, Volume 7: Language Testing and Assessment*. New York: Springer, 241-254.
- Thaiss, C.J. & Zawacki, T.M. (2006). *Engaged Writers, Dynamic Disciplines: Research on the Academic Writing Life*. Portsmouth/NH: Boynton/Cook.
- Wulff, S. & Gries, S.Th. (2011). A multifactorial study of the genitive alternation in L2 English. Paper presented at *The Second Tri-Annual Conference of the International Society for the Linguistics of English*, Boston University, 17-21 June, 2011.
- Wulff, S. & Römer, U. (2009). Becoming a proficient academic writer: Shifting lexical preferences in the use of the progressive. *Corpora* 4(2), 115-133.
- Zaytseva, E. (2011). Register, genre, rhetorical functions: Variation in English native-speaker and learner writing. In H. Hedeland, T. Schmidt & K. Wörner (eds) *Multilingual Resources and Multilingual Applications (Hamburg Working Papers in Multilingualism B 96)*, 239-242.

