

Research based on CALE focuses on the study of patterns and determinants of **lexico-grammatical variation** in **Advanced Learner Varieties (ALVs)**.

What are Advanced Learner Varieties?

Advanced learners of various language backgrounds have similar problems and face similar challenges on their way to near-native proficiency. For example, they typically struggle with the acquisition of **optional and/or highly L2-specific linguistic phenomena**. As for academic writing, many of their observed difficulties are due to a lack of understanding of **conventions of academic writing**, or a lack of practice, but not necessarily a result of interference from L1 academic conventions. In view of these similarities, we refer to the interlanguage of these learners as **Advanced Learner Varieties (ALVs)**.

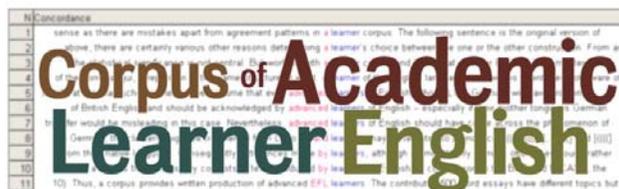
Why lexico-grammatical variation?

Since the late 1990s SLA research has seen an increasing interest in advanced stages of acquisition and questions of near-native competence. In research on **variation in interlanguage**, there has been a shift from the study of norm violations to an investigation of differential knowledge as evidence of conscious awareness of linguistic variation. This also marked a shift in focus from beginning/intermediate to advanced learners.

In our research, we examine L2 learners' acquisition of the influence of individual (or the interplay of several) factors that determine **constituent order** and the **choice of constructional variants** (such as genitive/dative alternation, verb-particle placement, focus constructions). Such variation has not been well researched in an SLA context to date.

Applications

The corpus and the research findings based on the corpus data will help us to develop a text-centred procedure to **assess advanced proficiency** based on linguistic criteria. In addition, the data will be used to develop **teaching materials** and to make recommendations for **teaching English for Academic Purposes**.



We are collecting texts and bio data from university students of English, mostly in the discipline of (applied) linguistics and related language studies (e.g. translation). In the long run, CALE may also include academic written English in **other disciplines**.

We welcome project partners who are **interested in joining the project** and would like to contribute texts from advanced learners of **other L1-backgrounds** (e.g. Polish, Portuguese, Swedish, Finnish etc.). Please approach us personally or send an e-mail to the principal investigator.

For further information, please contact:

Prof. Dr. Marcus Callies
Universität Bremen
FB 10 – Sprach- und Literaturwissenschaften
Bibliothekstraße, Gebäude GW 2, A 3400
D-28359 Bremen, Germany

@ callies@uni-bremen.de

Visit our Website! 

Our website informs about recent news and developments and gives an overview of recent and upcoming project-related presentations and publications:

<http://www.advanced-learner-varieties.info>

CALE

Corpus of Academic Learner English



 **Universität Bremen**

www.advanced-learner-varieties.info

What is the CALE?

The *Corpus of Academic Learner English* (CALE) is an electronic text corpus currently being compiled to represent written academic English produced by advanced learners of English as a foreign language ("academic learner writing"). CALE is thus what may be called a *Language for Specific Purposes* learner corpus, because it contains discipline- and genre-specific texts.



We are collecting texts and bio data from L2 learner of English who volunteer to have their texts included in the corpus. Most texts are produced as assignments in university courses.

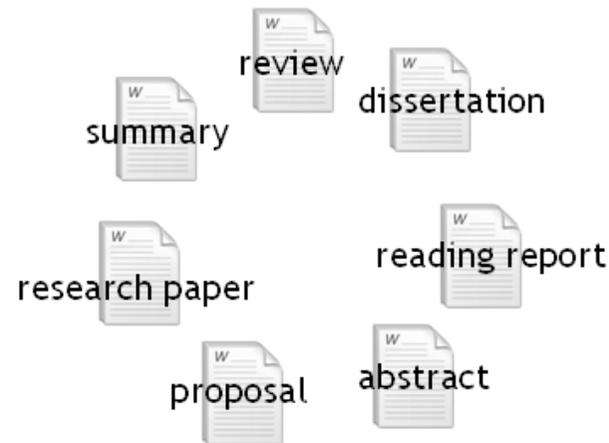
Why a new learner corpus?

Existing learner corpora like e.g. the *International Corpus of Learner English* largely include argumentative or literary writing of a general nature, not academic writing in a narrow sense. Therefore, various lexical, phraseological and syntactic phenomena that are typical of the academic register and important to study in the context of English for Academic Purposes are not represented (frequently enough) in these corpora.

Corpus Design and Composition

What type of material does the corpus include?

CALE is designed to include **seven different text types** ("genres") that are characteristic of **written academic discourse**:



The text classification developed for CALE is generally comparable with existing native-speaker control corpora. We have created clear textual profiles, adopting situational characteristics and linguistic features that have been identified for academic prose (see e.g. D. Biber & S. Conrad [2009], *Register, Genre, and Style*. Cambridge: CUP).

In our classification, the **communicative purpose/goal** of a text serves as the main classifying principle, and the seven genres are set apart in terms of four characteristic features:

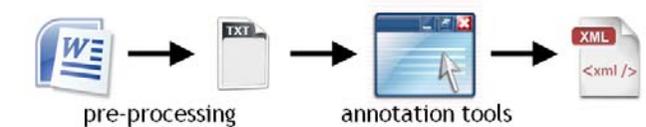
- a text's general purpose
- its specific purpose(s)
- the skills the author demonstrates
- the author's stance.

In addition, we list the major features of each text type as to

- structural features
- length
- functional features.

Corpus Annotation

Students submit their texts in electronic form. 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. With the help of **annotation tools** the body of the text is then annotated for **meta-textual**, i.e. **underlying structural features** (section titles, paragraphs, quotations, examples etc.).



Additionally, the texts are annotated for **metadata**, i.e. **learner variables** such as L1, age, gender, etc. and a set of **further 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.

Individual case studies based on the corpus will implement other levels of annotation, most importantly those adopting a **function-to-form approach** to the analysis of learner language such as the annotation of rhetorical functions (e.g. contrast, exemplification) and the lexico-grammatical means to express them.

Function-oriented levels of annotation, though still rare in learner corpus research, present researchers with a valuable opportunity to view learners as active language users, rather than demonstrating deficient knowledge of the target language. The advantage of a function-driven annotation is that it makes it possible to identify and pin down a near-complete inventory of lexical-grammatical means used to express various functional rhetorical functions in written discourse.