Corpora and corpus linguistic tools and methods are frequently used in the study of second language (L2) learning, most notably in Learner Corpus Research (LCR). LCR has contributed significantly to the description of interlanguages and many of its findings have resulted in useful applications for foreign language teaching and learning. Learner- and native-speaker corpora have also received increasing attention in the area of language testing and assessment (LTA; Barker 2010; Taylor & Barker 2008). Practical applications of corpora in LTA can range from corpus-informed to corpus-based and corpus-driven approaches, depending on how corpus data are actually put into practice, the aims and outcomes for LTA, and the degree of involvement of the researcher in the process of data retrieval, analysis and interpretation (Barker 2010; Callies, Zaytseva & Diez-Bedmar to appear).

More recently, researchers have also turned to corpora to inform, validate, and develop the way proficiency is operationalized in the Common European Framework of Reference for Languages (CEFR; Council of Europe 2001, 2009). While the CEFR has been highly influential in language testing and assessment, the way it defines proficiency levels using “can-do-statements” has been criticized, because they are often too impressionistic. For example, a learner at the C2 level is expected to maintain “consistent grammatical control of complex language”, whereas at C1 he/she should “consistently maintain a high degree of grammatical accuracy” (Council of Europe 2001, 2009). Such global, vague and underspecified descriptions have limited practical value to distinguish between proficiency levels and also fail to give in-depth linguistic details regarding individual languages or learners’ skills in specific registers. These shortcomings have led to an increasing awareness among researchers of the need to identify more specific linguistic descriptors or ‘criterial features’ which can be quantified by learner data. The aim of such corpus-based approaches is to add “grammatical and lexical details of English to CEFR’s functional characterisation of the different levels” (Hawkins & Filipovic 2012: 5).

While (learner) corpora have the potential to increase transparency, consistency and comparability in the assessment of L2 proficiency, several problems and challenges may also be encountered. One major difficulty is that “proficiency level” has often been a fuzzy variable in learner corpus compilation and analysis (Carlsen 2012), because, due to practical constraints, proficiency has mostly been operationalized and assessed globally by means of external criteria, typically learner-centred methods such as learners’ institutional status. However, recent studies show that global proficiency measures based on external criteria alone are not reliable indicators of proficiency for corpus compilation (Mukherjee 2009; Callies to appear 2013), and “hidden” differences in proficiency (e.g. Pendar & Chapelle 2008) often go undetected or tend to be disregarded in learner corpus analysis (e.g. Götz 2013). Thus, the field still seems to be in need of a corpus-based description of language proficiency to account for inter-learner variability and seek homogeneity in learner corpus compilation and L2 assessment. Another issue that has been
intensively debated is the appropriate basis of comparison for learner corpus data, i.e. against what yardstick learner performance should be compared and evaluated.

The aim of this workshop is to discuss the benefits in terms of current practices and developments, but also the challenges and possible obstacles of using both native-speaker reference corpora and learner corpora for testing and assessing L2 proficiency. We thus invite submissions that provide case studies exemplifying how corpora can be used for the assessment of L2 proficiency in both speaking and writing. In particular, submissions should address one of the following topics:

- **corpus compilation** (types of corpus data and their usefulness for testing purposes; proficiency as a fuzzy variable in learner corpus compilation and analysis; homogeneity vs. variability in corpus composition)

- **corpus comparability** (e.g. as to register/genre or task setting and conditions, i.e. testing vs. non-testing contexts, prompt, timing, access to reference works)

- the operationalization of (types of) proficiency in corpus approaches to testing and assessment

- the use of corpora in data-driven approaches to the assessment of proficiency (e.g. using corpus data to validate or complement human rating as in studies based on error-tagged learner corpora, or using corpus data (partially) independently of human rating).

Abstracts are invited for this workshop and should be 400 to 500 words long (excluding references). They should be submitted by e-mail to callies@uni-bremen.de and Sandra.Goetz@anglistik.uni-giessen.de by 1st February 2013. Notification of acceptance will be sent out in late February 2013.

**References**


Callies, M. (to appear 2013), Advancing the research agenda of Interlanguage Pragmatics: The role of learner corpora. *Yearbook of Corpus Linguistics and Pragmatics*.


