

Marcus Callies

A corpus approach to operationalizing and assessing writing proficiency in the academic register: The case of reporting verbs

Abstract

Der *Gemeinsame Europäische Referenzrahmen für Sprachen* (GER) ist vielfach kritisiert worden, insbesondere im Hinblick auf die praktische Umsetzbarkeit seiner sprach- und registerneutralen Kann-Beschreibungen. Die im GER verankerte Beschreibung der Kompetenzniveaus bedarf daher einer Spezifikation in Bezug auf Einzelsprachen und spezifische Register, um einen zuverlässigen Maßstab für die praktische Bewertung von Sprachkompetenzen leisten zu können. In diesem Beitrag wird mittels einer Fallstudie ein Ansatz präsentiert, in dem Sprachkorpora eingesetzt werden, um einige der Einschränkungen des GER-basierten Testens und Prüfens von Kompetenzniveaus zu überwinden. Es werden dabei explizite, quantifizierbare und sprachspezifische Deskriptoren identifiziert und angewendet, mit deren Hilfe die Erhebung von Schreibkompetenzen im Wissenschaftsenglisch ermöglicht wird. Der spezifische Deskriptor, der hier exemplarisch untersucht wird, sind sog. Berichtsverben, die ein charakteristisches Merkmal wissenschaftlicher Texte sind.

1. Introduction: Learner corpora in Language Testing and Assessment

Learner corpora are systematic collection of authentic, continuous and contextualized language use produced by foreign/second language (L2) learners. In contrast to (quasi-)experimental, elicited data used traditionally in second language acquisition (SLA) research, learner corpora comprise large amounts of data produced in rather uncontrolled and open-ended speaking or writing tasks (e.g. open interviews or essays), and are stored in electronic format to be processed with the help of corpus software tools. Learner corpora have been compiled and used for almost two decades, giving rise to the field of Learner Corpus Research (LCR; see Granger/Meunier/Gilquin 2015). LCR serves to provide an in-depth description of interlanguages, to give SLA theories a more solid empirical foundation, and to help produce tools and teaching materials designed for the needs of specific learner populations.

More recently, learner corpora have also been used to inform and complement the testing and assessment (LTA) of L2 proficiency (see Callies/Götz 2015). Learner corpora are increasingly used in LTA (Taylor/Barker 2008, Barker 2010) and have the potential to increase transparency, consistency and comparability in the assessment of L2 proficiency. LCR thus responds to some of the limitations of more established ways of assessing proficiency in LTA where writing proficiency has largely been assessed by expert raters' scale-based evaluation of the written sections of standardised language tests. These test results are then usually evaluated and interpreted in terms of the "degree of task fulfilment and evidence of target language control according to criteria such as *communicative effectiveness, register, organisation, linguistic range and accuracy*" (Hawkey/Barker 2004: 122f; emphasis in original). Learners are then assigned to a proficiency level in accordance with expectations of what L2 learners can do at a particular level as specified, for example, in the *Common European Framework of Reference for Languages* (CEFR). However, proficiency assessment by means of global rating scales, even when carried out by trained and experienced expert raters, is influenced by the subjectivity and variability of human rating.

More recently, therefore, a new research strand has proposed to make use of quantifiable linguistic descriptors and learner corpus data to inform, validate, and advance the assessment of L2 proficiency (e.g. Taylor/Barker 2008; Hawkins/Buttery 2009, 2010; Barker 2010, 2013; Hawkins/Filipovic 2012). Realizing that the can-do statements used to describe and define proficiency levels in the CEFR do not provide lan-

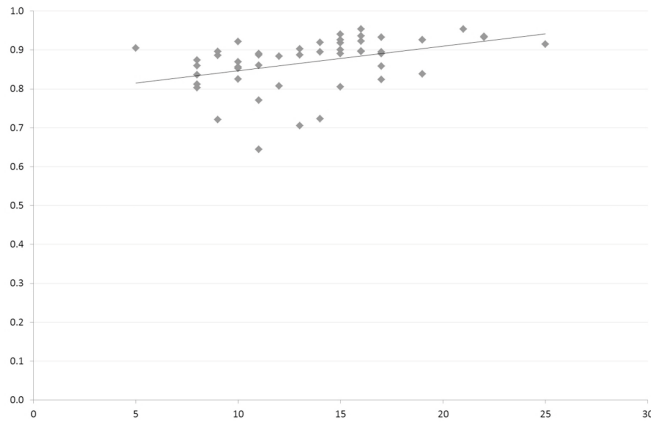


Fig. 5: Verb types correlated with D ($r = 0.40$)

4. Conclusion

It has been argued in this paper that learner corpora can inform, complement and possibly advance the assessment of L2 proficiency vis-à-vis the CEFR. The identification and operationalisation of a set of linguistic descriptors that are a) language- and register-specific, b) quantifiable and c) subject to (semi-) automatic processing can inform a data-driven assessment of writing proficiency in the academic register involving three steps:

- selection of linguistic features for academic prose in terms of their keyness, operationalizability and “late acquisition” (corpus-informed)
- retrieval of descriptors from corpus (corpus-based)
- classification and assessment of proficiency using statistical techniques (corpus-driven).

While the corpus-driven approach has not yet been widely used in LTA, it seems particularly promising and useful for a corpus approach to assessing L2 proficiency based on linguistic descriptors, such as those that are typical of a specific register, e.g. academic writing.

References

- ANTHONY, L. (2015): TagAnt (Version 1.2.0). Tokyo: Waseda University.
<http://www.laurenceanthony.net/>

- ANTHONY, L. (2016): AntConc (Version 3.4.4). Tokyo: Waseda University. <http://www.laurenceanthony.net/>
- BARKER, F. (2010): How can corpora be used in language testing? In: O'Keeffe, A./McCarthy, M. (eds.): *The Routledge Handbook of Corpus Linguistics*. New York: Routledge, pp. 633–645.
- BARKER, F. (2013): Assessment and testing: Overview. In: Chapelle, C. A. (ed.): *The Encyclopedia of Applied Linguistics*. New York: Blackwell, pp. 1360–1366.
- BIBER, D./CONRAD, S. (2009): *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- CALLIES, M. (2013): Agentivity as a determinant of lexico-grammatical variation in L2 academic writing. In: *International Journal of Corpus Linguistics*, 18:3, pp. 357–390.
- CALLIES, M./GÖTZ, S. (2015): Learner corpora in language testing and assessment: Prospects and challenges. In: Callies, M./Götz, S. (eds.), *Learner Corpora in Language Testing and Assessment*. Amsterdam: Benjamins, pp. 1–9.
- 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. In: *Dutch Journal of Applied Linguistics*, 2:1, pp. 126–132.
- CALLIES, M./DIEZ-BEDMAR, M. B./ZAYTSEVA, E. (2014): Using learner corpora for testing and assessing L2 proficiency. In: Leclercq, P./Hilton, H./Edmonds, A. (eds.): *Proficiency Assessment Issues in SLA Research: Measures and Practices*. Clevedon: Multilingual Matters, pp. 71–90.
- COXHEAD, A. (2000): A new academic word list. In: *TESOL Quarterly*, 34, pp. 213–238.
- GARDNER, D./DAVIES, M. (2014): A new academic vocabulary list. In: *Applied Linguistics*, 35:3, pp. 305–327.
- GRANGER, S./PAQUOT, M. (2009): Lexical verbs in academic discourse: A corpus-driven study of learner use. In: Charles, M./Pecorari, D./Hunston, S. (eds.): *Academic writing. At the interface of corpus and discourse*. London: Continuum, pp. 193–214.
- GRANGER, S./MEUNIER, F./GILQUIN, G. (eds.) (2015): *Cambridge Handbook of Learner Corpus Research*. Cambridge: CUP.
- HAWKEY, R./BARKER, F. (2004): Developing a common scale for the assessment of writing. In: *Assessing Writing*, 9, pp. 122–159.
- HAWKINS, J./BUTTERY, P. (2009): Using learner language from corpora to profile levels of proficiency: Insights from the English Profile Programme. In: Taylor, L./Weir, C. J. (eds.): *Language Testing Matters*:

- Investigating the Wider Social and Educational Impact of Assessment. Cambridge: Cambridge University Press, pp. 158–175.
- HAWKINS, J./BUTTERY, P. (2010): Criterial features in learner corpora: Theory and illustrations. In: *English Profile Journal*, 1:1, pp. 1–23.
- HAWKINS, J./FILIPOVIC, L. (2012): *Criterial Features in L2 English*. Cambridge: Cambridge University Press.
- HYLAND, K. (1999): Academic attribution: citation and the construction of disciplinary knowledge. In: *Applied Linguistics*, 20 (3), pp. 341–367.
- HYLAND, K. (2002): Activity and evaluation: Reporting practices in academic writing. In: Flowerdew, J. (ed.): *Academic Discourse*. Harlow: Longman, pp. 115–130.
- JARVIS, S. (2013): Capturing the diversity in lexical diversity. In: *Language Learning* 63: Suppl. 1, pp. 87–106.
- NEFF, J./DAFOUZ, E./HERRERA, H./MARTINEZ, F./RICA, J. P./DIEZ, M./PRIETO, R./SANCHO, C. (2003): Contrasting learner corpora: The use of modal and reporting verbs in the expression of writer stance. In: Granger, S./Petch-Tyson, S. (eds.): *Extending the Scope of Corpus-Based Research. New Applications, New Challenges*. Amsterdam: Rodopi, pp. 211–230.
- ORTEGA, L./BYRNES, H. (2008): The longitudinal study of advanced L2 capacities: An introduction. In: Ortega, L./Byrnes, H. (eds): *The Longitudinal Study of Advanced L2 Capacities*. New York: Routledge/Taylor & Francis, pp. 3–20.
- PAQUOT, M. (2010): *Academic Vocabulary in Learner Writing*. London: Continuum.
- TAYLOR, L./BARKER, F. (2008): Using corpora for language assessment. In: Shohamy, E./Hornberger, N. H. (eds.): *Encyclopedia of Language and Education*. 2nd Edition, Volume 7: Language testing and assessment. New York: Springer, pp. 241–254.
- THOMAS, S./HAWES, T. P. (1994): Reporting verbs in medical journal articles. In: *English for Specific Purposes*, 13 (2), pp. 129–148.
- WULFF, S./GRIES, S. (2011): Corpus-driven methods for assessing accuracy in learner production. In: Robinson, P.(ed.): *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*. Amsterdam: Benjamins, pp. 61–87.

Kontakt Daten des Autors

Marcus Callies, University of Bremen,
callies@uni-bremen.de