The influence of L1 Dutch on cohesion in L2 German writing: Results from a contrastive corpus-based analysis of L1 and L2 students´ writing in German

Second language (L2) writers struggle with cohesion (Crossley & McNamara, 2012). One of the reasons is that they tend to rely on native language (L1) strategies to create cohesive texts (Breindl, 2016) which may differ from the strategy used in the L2 (e.g., Pit, 2007). This challenge has been documented in studies into L2 English (e.g., Appel & Szeib, 2018 Johnson, 2017; Stemmer, 1991). For example, Appel and Szeib (2018) investigated the influence of L1 Arab, L1 Chinese and L1 French on linking adverbials in L2 English and found differences in the use specific to the L1s, such as an overuse of appositional linking by L1 French writers and an overuse of contrastive linking by L1 Chinese writers. However, in stark contrast to the burgeoning research in cohesion in L2 English (e.g., Crossley & McNamara, 2011, 2012; Crossley et al., 2019; Das et al., 2017) research into cohesion in L2 German has been scarce to date, with only a handful of studies into texts produced by writers with heterogeneous L1 backgrounds. These contrastive studies focus on, for example, the different uses of connectives (Walter, 2007; Walter & Schmidt, 2008) and pronominal adverbs (Belz, 2005; Breindl, 2016; Strobl, 2020a) in L1 and L2 writing.

Our study aims fill this gap by furthering research into contrastive studies of connectives in L1 and L2 German writing, focusing on L2 writers with a homogeneous L1 language background (i.e., Dutch). The contrastive analysis will be based on the Belgisches Deutschkorpus (Beldeko) (Strobl, 2020b) and the German Summary Corpus (GerSumCo). These are two corpora that have recently been compiled to analyse cohesion in L2 language. Beldeko consists of 301 texts written by advanced students of L2 German in an academic writing course. GerSumCo is still growing and to date includes 47 texts written by L1 German students. The texts are summaries (of the same source texts) that were produced under comparable conditions. The corpora have been pre-processed and automatically annotated with part-of-speech tags and lemmas. Additionally, the connectives were automatically pre-annotated using DimLex (Scheffler & Stede, 2016; Stede, 2002), a database containing German connectives and their corresponding PDTB3 tags (Webber et al., 2019). After automated pre-annotation, the data were manually corrected and enriched using the online annotation platform Inception (Klie et al., 2018). The three trained annotators used our own guidelines for the annotation of connectives, which are based on PDTB3.

The preliminary analysis of the corpora via R revealed a higher density of connective use in L2 summaries than in L1 summaries. In terms of semantic types, however, we can see a similar distribution. Crossley and McNamara (2012) documented a similar overuse of connectives in L2 English writing by students with a lower language proficiency, in comparison with L2 English writing by students with a higher language proficiency. They found that highly proficient writers tend to rely more on implicit cohesion (e.g., semantic overlap) than on explicit cohesion (e.g., connectives) which they called the "reverse cohesion effect". In our presentation, we will present the results of our first contrastive analysis of connective use in L1 and L2 German, discussing patterns in light of contrastive research on German and Dutch connectives (Pit, 2007) and research on connectives use in L1 German (e.g., Kunz et al., 2021).

References

Appel, R., & Szeib, A. (2018). Linking adverbials in L2 English academic writing: L1-related differences. *System*, *78*, 115–129. https://doi.org/10.1016/j.system.2018.08.008

Belz, J. A. (2005). Corpus-driven characterizations of pronominal da-compound use by learners and native speakers of German. *Die Unterrichtspraxis/Teaching German*, *38*(1), 44–60. https://doi.org/10.1111/j.1756-1221.2005.tb00041.x

Breindl, E. (2016). Konnexion in argumentativen Texten. Gebrauchsunterschiede in Deutsch als L2 vs. Deutsch als L1. *Normalität in der Sprache* (pp. 37–64). Buske.

Crossley, S., & McNamara, D. (2011). Shared features of L2 writing: Intergroup homogeneity and text classification. *Journal of Second Language Writing, 20*(4), 271–285. https://doi.org/10.1016/j.jslw.2011.05.007

Crossley, S. & McNamara, D. (2012). Predicting second language writing proficiency. The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, *35*(2), 115–135. https://doi.org/10.1111/j.1467-9817.2010.01449.x

Crossley, S. A., Kyle, K., & Dascalu, M. (2019). The tool for the automatic analysis of cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods, 51*(1), 14–27. https://doi.org/10.3758/s13428-018-1142-4

Das, D., Scheffler, T., Bourgonje, P., & Stede, M. (2018). Constructing a lexicon of English discourse connectives. *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue* (pp. 360–365). Association for Computational Linguistics. https://doi.org/10.18653/v1/W18- 5042

Johnson, M. (2017). Improving cohesion in L2 writing: A three-strand approach to building lexical cohesion. *English Teaching Forum*, *55*, 2–13.

Klie, J. C., Bugert, M., Boullosa, B., de Castilho, R. E., & Gurevych, I. (2018). The inception platform: Machine-assisted and knowledge-oriented interactive annotation. *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrationsm* (pp. 5–9). Association for Computational Linguistics.

Kunz, K., Lapshinova-Koltunski, E., Martínez Martínez, J., Menzel, K., & Steiner, E. (2021). *GECCo - German–English Contrasts in Cohesion: Insights from corpus-based studies of languages, registers and modes*. De Gruyter Mouton. https://doi.org/10.1515/9783110711073

Pit, M. (2007). Cross-linguistic analyses of backward causal connectives in Dutch, German and French. *Languages in Contrast*, *7*(1), 53–82. https://doi.org/10.1075/lic.7.1.04pit

Scheffler, T., & Stede, M. (2016). Adding semantic relations to a large-coverage connective lexicon of German. *Proceedings of LREC* (pp. 1008–1013). ELRA.

Stede, M. (2002). DiMLex: A lexical approach to discourse markers. In A. Lenci & V. Di Tomaso (Eds.), *Exploring the lexicon: Theory and computation* (pp. 1–15). Edizioni dell'Orso.

Stemmer, B. (1991). *Kohäsion im gesprochenen Diskurs deutscher Lerner des Englischen*. J. Groos.

Strobl, C. (2020a). Darum sind Pronominaladverbien eine Herausforderung für Deutschlerner: Eine korpusbasierte kontrastive Interimssprachenanalyse hierzu. *Germanistische Mitteilungen*, *45*(1), 89–111.

Strobl, C. (2020b) *Beldeko Summary Corpus v1.0.0*. Eurac Research CLARIN Centre. http://hdl.handle.net/20.500.12124/15

Walter, M. (2007). Hier wird die Wahl schwer, aber entscheidend: Konnektorenkontraste im Deutschen. In H.-J. Krumm (Ed.), *Theorie und Praxis: Österreichische Beiträge zu Deutsch als Fremdsprache* (pp. 145–161). StudienVerlag.

Walter, M., & Schmidt, K. (2008): "Und das ist auch gut so": Der Gebrauch des satzinitialen und bei fortgeschrittenen Lernern des Deutschen als Fremdsprache. In B. Ahrenholz, U. Bredel, W. Klein, M. Rost-Roth, & R. Skiba (Eds.), *Empirische Forschung und Theoriebildung. Beiträge aus Soziolinguistik, Gesprochene-Sprache- und Zweitspracherwerbsforschung: Festschrift für Norbert Dittmar zum 65. Geburtstag* (pp. 331–342). Peter Lang.

Webber, B., Prasad, R., Lee, A., & Joshi, A. (2019). The Penn Discourse Treebank 3.0 Annotation Manual. *University of Pennsylvania*.