Reading demands and language proficiency in foreign language learning: An exploration of linguistic complexity modeling

Zarah Weiss and Detmar Meurers

The language input available to language learners is considered to be a driving force for Second Language Acquisition (SLA). Input that is just above the level of the learner is assumed to best foster learning, which depending on the SLA tradition is characterized as $i + 1$ input of Krashen (1981), input in the Zone of Proximal Development in socio-cultural approaches (Lantolf, Thorne, & Poehner, 2015), or L2 development reflecting input in usage-based SLA approaches (Ellis & Collins, 2009). Yet how can one make the envisaged adaptive relation between the proficiency of a given learner and the input concrete and useful for education in practice?

We explore this relation based on the notion of linguistic complexity. As part of the triad Complexity, Accuracy, and Fluency it is used in SLA to characterize language proficiency based on analyses of learner productions (cf., e.g., Housen & Kuiken, 2009), and such SLA complexity measures have also been shown to be effective for characterizing the complexity of reading material (cf., e.g., Vajjala & Meurers, 2012). Chen and Meurers (2019) therefore proposed a Computer-Assisted Language Learning tool using high-dimensional vectors of linguistic complexity measures as a fine-grained link between input and output of learners. While conceptually this appears attractive, how empirically viable is this linguistic complexity link?

In this presentation, we explore this question by analyzing and comparing the linguistic complexity of graded reading materials with that of CEFR-rated learner productions. We focus on German as a Foreign Language to help extend the focus and relevance of research in this domain to languages beyond the dominant English. For the graded reading material, we compiled a corpus of 1.446 articles published for beginner, intermediate, and advanced learners of German between 2018 and 2019 in 45 magazines by a leading European publisher for foreign language reading materials. For the learner productions, we make use of the MERLIN corpus (Wisniewski et al., 2013) of learner writings for official standardized foreign language certification tests. Building on our linguistic complexity analysis of MERLIN in Weiss and Meurers (2019), we use the same approach to obtain a broad linguistic complexity analysis of the new graded reader corpus and compare the results for the two corpora at the complexity feature level as well at the document level using supervised machine learning classification experiments. After discussing classification accuracies and insights about the most informative features for within-corpus classification experiments, we turn to cross-corpus analyses designed to probe into the link between input and output complexity in this foreign language learning context.

In sum, the presentation will explore the use of linguistic complexity analysis as a link between learner input and their output, present novel complexity analyses on a new German as a Foreign Language reading corpus, and relate the analysis to that of a graded learner corpus. On the practical side, the approach supports the development of tools for the retrieval of individually adaptive reading materials for foreign language teaching and learning.

References

Chen, X., & Meurers, D. (2019). Linking text readability and learner proficiency using linguistic complexity feature vector distance. *Computer-Assisted Language Learning*. (https://doi.org/10.1080/09588221.2018.1527358)

Ellis, N., & Collins, L. (2009). Input and second language acquisition: The roles of frequency, form, and function introduction to the special issue. *The Modern Language Journal*, *93*(3), 329-335. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-4781.2009.00893.x doi: 10.1111/j.1540-4781.2009.00893.x

Housen, A., & Kuiken, F. (2009). Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 461–473.

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Prentice Hall International. Retrieved from http://www.sdkrashen.com/SL_Acquisition_and_Learning/index.html

Lantolf, J. P., Thorne, S. L., & Poehner, M. E. (2015). Sociocultural theory and second language development. In *Theories in second language acquisition: An introduction.* Routledge New York, NY.

Vajjala, S., & Meurers, D. (2012). On improving the accuracy of readability classification using insights from second language acquisition. In (pp. 163–173). (http://aclweb.org/anthology/W12-2019.pdf)

Weiss, Z., & Meurers, D. (2019). Broad linguistic modeling is beneficial for German L2 proficiency assessment. In A. Abel, A. Glaznieks, V. Lyding, & L. Nicolas (Eds.), *Widening the scope of learner corpus research. selected papers from the fourth learner corpus research conference.* Louvain-La-Neuve: Presses Universitaires de Louvain.

Wisniewski, K., Schöne, K., Nicolas, L., Vettori, C., Boyd, A., Meurers, D., ... Hana, J. (2013). Merlin: An online trilingual learner corpus empirically grounding the european reference levels in authentic learner data. In *Ict for language learning 2013, conference proceedings, florence, italy. libreriauniversitaria. it edizioni.*