



Advances and challenges in automatic learner language annotation for a deepened understanding of variation inside and across developmental stages

Matthias Schwendemann, Annette Portmann, Josef Ruppenhofer, Torsten Zesch, Katrin Wisniewski

The idea of developmental stages has played a major role in SLA research, particularly in Processability Theory (PT, Pienemann, 1998, 2005), where variation is conceived of in a rather narrow way (as constrained by processability). Recently, researchers both from inside the PT paradigm (e.g., Nicholas et al., 2019) and researchers with other backgrounds have emphasized the need to take a broader perspective on variation inside and across developmental stages. On a methodological level, however, an obstacle for variation research has been the limited size of corpora most SLA studies on developmental stages have used so far.

This is the background of the DAKODA project (*Data literacy in German as a Foreign/Second Language: Exploring computational linguistic approaches to the analysis of German L2 developmental stages*). DAKODA aims to explore the usefulness of language technology for the automated analysis of developmental stages in L2 German. To that aim, we collect, consolidate and make available a large number of German learner corpora in a repository and via an online interface. For the first time, it will be possible to access these corpora from a common platform. We use foreign language corpora (e.g., Falko, Beldeko), purpose-built SLA corpora (e.g., ESF, ZISA corpus data), and yet unpublished corpora (e.g., MULTILIT). The language samples thus vary in terms of learners' background variables, elicitation procedures, and other features. If the automatic analyses are successful, the multitude of heterogeneous data in DAKODA will make it possible to take a closer look at inter- and intra-individual variation within and across developmental stages and potentially allow for a re-evaluation of their presumed universality.

In our presentation, we will outline the DAKODA project and offer some first insights into the possibilities and challenges of (automatic) analysis of developmental stages in learner language.

References

Nicholas, H., Lenzing, A. & Roos, J. (2019). Chapter 17. How does PT's view of acquisition relate to the challenge of widening perspectives on SLA? In A. Lenzing, H. Nicholas & J. Roos (Hrsg.), *Widening Contexts for Processability Theory* (Bd. 7, S. 391–398). John Benjamins Publishing Company. <https://doi.org/10.1075/palart.7.17nic>

Pienemann, M. (1998). *Language Processing and Second Language Development* (Bd. 15). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.15>

Processability Theory. In M. Pienemann (Hrsg.), *Studies in bilingualism. Cross-Linguistic Aspects of Processability Theory* (Bd. 30, S. 1–60). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.30.03pie>