

Explorations of Variability. Evidence from L1 and L2 Corpora of German

Anke Lüdeling

Humboldt-Universität zu Berlin

In recent years, corpus linguistics has learned a lot about variability between corpora and texts. We have seen research on external factors such as task effects (Crowther et al. 2015, Gablasova et al. 2017, Schnur & Rubio 2021, Weiss 2017) and setting, as well as speaker factors such as age, socio-economic properties, aptitude, or motivation (Birdsong 2018, Dörnyei & Ryan 2015, Granena 2013, Larsen-Freeman 2018, among many others).

Building on this research, this paper will dig even further into different aspects of variability and discuss a number of theoretical and methodological implications. Using two German corpora that are maximally controlled for external factors with matching L1 and L2 subcorpora (Kobalt, Zinsmeister et al. 2012 and Falko, Reznicek et al. 2012) and deeply annotated for syntactic categories as well as morphological subclasses (see Lukassek *et al.* 2022, Shadrova 2021), this paper will investigate intra- and inter-speaker variation with respect to morphology and syntax. The results of recent research I have conducted with colleagues indicate a surprisingly high degree of variance between L1 speakers in the distribution of word formation subclasses (cf. Shadrova et al. 2021), which challenges the construct of native speaker homogeneity beyond stable, stratified, and situational variation and raises methodological questions for comparative L1-L2 studies as well as to the role of frequency in entrenchment.

I will further present evidence for intra-individual differences, looking into procedural factors, such as priming and self-priming and within-text register fluctuation, highlighting the necessity of accounting for text dynamics and aspects of (the acquisition of) register knowledge. At the same time, the distribution of syntactic categories such as dependencies and parts of speech is much less variable across speakers and even between corpora, suggesting categorical differences between syntax and the lexicon in production and challenge the notion of “constructions all the way down” (Goldberg 2006, 18).

Based on these explorations of variability between and within speakers, as well as between linguistic layers, I will discuss the potential of small, deeply annotated and well-understood corpora, which are ideally suited to accommodate to the needs of careful linguistic analysis in a complex space of interactions and path-dependencies.

References

- Birdsong, David (2018) Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology* 9(81), <https://doi.org/10.3389/fpsyg.2018.00081>.
- Crowther, Dustin, Pavel Trofimovich, Talia Isaacs & Kazuya Saito (2015) Does a Speaking Task Affect Second Language Comprehensibility? *The Modern Language Journal* 99(1), 80-95.
- Dörnyei, Zoltán & Stephen Ryan (2015) *The Psychology of the Language Learner Revisited* Routledge.
- Gablasova, Dana, Vaclav Brezina, Tony McEnery & Elaine Boyd (2017) Epistemic stance in spoken L2 English: The effect of task and speaker style. *Applied Linguistics* 38(5), 613-637.
- Goldberg, Adele E. (2016) *Constructions at Work: The Nature of Generalizations in Language*. Oxford University Press.
- Granena, Gisela (2013) Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning* 63(4), 665-703.
- Larsen-Freeman, Diane (2018) Looking ahead: Future directions in, and future research into, second language acquisition. *Foreign Language Annals* 51(1), 55-72.

- Lukassek, Julia, Roodabeh Akbari & Anke Lüdeling (2022) Guidelines for the morphological annotation of nouns in Falko. To appear in *REALIS White Paper Series* <https://sfb1412.hu-berlin.de/>.
- Reznicek, Marc, Maik Walter, Kari Schmidt, Anke Lüdeling, Hagen Hirschmann, Cedric Krummes & Torsten Andreas (2012) Das Falko-Handbuch: Korpusaufbau und Annotationen. Unpublished, <https://hu-berlin.de/falko>.
- Schnur, Erin & Fernando Rubio (2021) Lexical complexity, writing proficiency, and task effects in Spanish Dual Language Immersion. *Language Learning and Technology* 25(1),53-72. <https://hdl.handle.net/10125/73425>.
- Shadrova, Anna (2021) Kobalt: Extension Corpus and Annotation Guidelines for Verb Classification and Dependency Adjustments. Zenodo. <https://doi.org/10.5281/zenodo.5730224>.
- Shadrova, Anna, Pia Linscheid, Julia Lukassek, Anke Lüdeling & Sarah Schneider (2021) A challenge for contrastive L1/L2 corpus studies: Large inter-and intra-individual variation across morphological, but not global syntactic categories in task-based corpus data of a homogeneous L1 German group. *Frontiers in Psychology* 12, <https://doi.org/10.3389/fpsyg.2021.716485>.
- Weiss, Zarah (2017) Using measures of linguistic complexity to assess German L2 proficiency in learner corpora under consideration of task-effects. Unpublished MA thesis. www.sfs.uni-tuebingen.de/zweiss.
- Zinsmeister, Heike, Marc Reznicek, Julia Ricard-Brede, Christina Rosén & Dirk Skiba (2012) Das Wissenschaftliche Netzwerk „Kobalt-DaF “ Korpusbasierte Analyse von Lernertexten für Deutsch als Fremdsprache. *Zeitschrift für germanistische Linguistik* 40(3), 457-458.