

## Assessing compatibility and stability of individual grammars through replication

Gradient variability is studied prominently in Probabilistic Grammar (Bresnan 2007). A major strain of research in Probabilistic Grammar focuses on cases where speakers chose between two or more superficially interchangeable options (*alternations*) and the choices are controlled probabilistically by contextual features and soft constraints (e. g., Bresnan 2007, Gries 2017). Often, corpus evidence is used, but some studies also uses experimental data either exclusively or in order to corroborate the corpus findings. Some studies show how corpus data and experimental data converge (Bresnan et al. 2007, Durrant & Doherty 2010, Gries & Wulff 2005, Gries et al. 2005), but a number of other studies led to diverging or only partially converging results between corpus and experimental evidence (Arppe & Järvikivi 2007, Dąbrowska 2014, Mollin 2009). More importantly, Dąbrowska (2008, 2012, 2015) presented evidence showing that individual speakers have partially incompatible grammars (not reducible to dia- or sociolectal variation). This poses some problems for the use of massively pooled data in corpora. In psycholinguistics, participant-level variables (such as age) are often used to model between-speaker variation (e. g., Huettig & Janse 2016), but Verhagen & Mos (2016) address the possibility that there might be random between-speaker variability in the sense of Dąbrowska (*incompatibility*), and individual speaker grammars might also be subject to random fluctuations (*instability*). There are thus at least six potential sources of variability to consider in alternation research: [1] soft context-sensitive semantic/syntactic constraints, [2] functional within-speaker variation (register), [3] incompatibility, [4] instability, [5] noise, including performance effects, [6] artefacts of the experiment.

My research presented here addresses [3] through [6]. I show results of two repeated replications of two previously published experiments (split-100 and self-paced reading [SPR]) on binary morphosyntactic alternations in German. In both cases, determining context-sensitive semantic and morphosyntactic constraints influencing the alternation was the objective of the original study, and the point of the replications was to find out whether and how incompatibility and instability mar the inferences about these constraints. The repeated replication consisted of two replications of the original studies with the same highly homogeneous groups of participants with two months in between, allowing for a detailed analysis of incompatibility and instability effects. Both original experiments were reported alongside parallel corpus studies, showing strong convergence (split-100) and weak but significant convergence (SPR) between corpus and experimental evidence. We thus have: [1] the corpus study, [2] the original experiment, [3] the first replication, [4] the second replication with the same participants as in [3]. The analysis shows incompatibility effects (like in Dąbrowska op. cit.) inasmuch as large separate groups of participants strongly prefer either variant A over B or variant B over A consistently in the split-100 task. The same effect seems to be observable in the SPR data with variant A or B (depending on the speaker) incurring a reading time delay regardless of other factors (further analysis pending). Turning to instability effects, for roughly two thirds of the participants, split-100 decisions are stable between the replications. The other third reacts more or less randomly across the two experiment runs. I argue that this is likely due to some problems with the split-100 task itself. More dramatically, over half of the participants in the SPR experiment show no stable behaviour in the two replications across all target stimuli, which means this cannot be due purely to performance effects. I discuss this w. r. t. the usability of SPR in alternation research. Also, I discuss the sensitivity of SPR to specifics of the experiment (cf. the wildly diverging SPR results from ten papers on Chinese relative clause processing in Vasisht 2015: 8). Importantly, however, the fundamental findings about the probabilistic semantic and morphosyntactic constraints controlling the alternation turn out to be robust across the corpus studies and all experiments. I discuss how this robustness might come about in the face of incompatibility and instability, and how this affects linguistic theory, experimental practice and corpus studies, as well as the statistical analysis of the findings (building on the discussion in Barr et al. 2013, Bates et al. 2015, Matuschek et al. 2017).

## References

- Arppe, Antti & Juhani Järvi­kivi. 2007. Every method counts: combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3(2), 131–159.
- Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language* 68(3), 255–278.
- Bates, Douglas M., Reinhold Kliegl, Shra­van Vasishth & R. Harald Baayen. 2015. *Parsimonious Mixed Models*.
- Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: linguistics in search of its evidential base* (Studies in Generative Grammar), 77–96. Berlin/New York: De Gruyter Mouton.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Koninklijke Nederlandse Akademie van Wetenschappen.
- Dąbrowska, Ewa. 2008. The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: an empirical test of usage-based approaches to morphology. *Journal of Memory and Language* 58, 931–951.
- Dąbrowska, Ewa. 2012. Different speakers, different grammars: individual differences in native language attainment. *Linguistic Approaches to Bilingualism* 2, 219–253.
- Dąbrowska, Ewa. 2014. Words that go together: measuring individual differences in native speakers' knowledge of collocations. *The Mental Lexicon* 9(3), 401–418.
- Dąbrowska, Ewa. 2015. Individual differences in grammatical knowledge. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of cognitive linguistics*, 650–668. De Gruyter.
- Durrant, Philip & Alice Doherty. 2010. Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory* 6(2), 125–155.
- Gries, Stefan Th. 2017. Syntactic alternation research. taking stock and some suggestions for the future. In Ludovic De Cuypere, Clara Vanderschueren & Gert De Sutter (eds.), *Current trends in analyzing syntactic variation*, vol. 31 (Belgian Journal of Linguistics), 7–27. Amsterdam: Benjamins.
- Gries, Stefan Th., Beate Hampe & Doris Schönefeld. 2005. Converging evidence: bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16(4), 635–676.
- Gries, Stefan Th. & Stefanie Wulff. 2005. Do foreign language learners also have constructions? evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics* 3, 182–200.
- Huettig, Falk & Esther Janse. 2016. Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience* 31(1), 80–93.
- Matuschek, Hannes, Reinhold Kliegl, Shra­van Vasishth, Harald Baayen & Douglas M. Bates. 2017. Balancing type I error and power in linear mixed models. *Journal of Memory and Language* 94, 305–315.
- Mollin, Sandra. 2009. Combining corpus linguistic and psychological data on word co-occurrences: corpus collocates versus word associations. *Corpus Linguistics and Linguistic Theory* 5(2), 175–200.
- Vasishth, Shra­van. 2015. *A meta-analysis of relative clause processing in Mandarin Chinese using bias modelling*. School of Mathematics & Statistics of the University of Sheffield dissertation.
- Verhagen, Véronique & Maria Mos. 2016. Stability of familiarity judgments: individual variation and the invariant bigger picture. *Cognitive Linguistics* 27(3), 307–344.