

Does UID constrain the usage and processing of Gapping in German?

Bozhidara Hristova, Robin Lemke, Lisa Schäfer, Heiner Drenhaus, Ingo Reich
Saarland University

We investigate the choice between Gapping (Ross, 1967) and a fully-fledged coordination (1), considering the Uniform Information Density (UID, Levy and Jaeger, 2007) principle. UID states that speakers prefer to distribute information (Shannon, 1948) uniformly across the utterance. We suggest that speakers use Gapping to omit predictable expressions and thus avoid information troughs. By doing so, they efficiently utilize the comprehenders' processing capacities (Fenk & Fenk, 1980). We focus on the omission of the inflected verb in the second conjunct (C2) of coordinations like (2) and modulate the predictability of the C2 verb through the number of objects in context (Object Number). We expect that mentioning two objects rather than one decreases the predictability of the C2 verb since readers expect the second object to be related to another action.

Experiment 1 (written sentence completion, N = 160) served two goals: (i) norming stimuli for follow-up experiments and (ii) testing whether Gapping is more frequent when the C2 verb is more likely. We manipulated the predictor Object Number (one/two) in German sentence pairs like (2) (32 item sets, 32 fillers). Target coordinations were cut off after the C2 subject (i.e., *Max*). We tested two hypotheses: H1) Subjects produce the same verb as the verb in the first conjunct (C1) more often with only one object in the context (main Object Number effect); and H2) Gapping is produced more often (than the nonelliptical form) when the C2 verb is more likely (main Verb Predictability effect). A GLMM (Bates et al., 2015) analysis¹ revealed a significant effect of Object Number ($z = 11.99, p < 0.001$): There were more identical (C2 verb = C1 verb) continuations in the one-object condition (Fig. 1), supporting H1. However, we found no significant Verb Predictability effect on the ratio of Gapping in identical continuations ($z = -0.42, p > 0.05$): In both conditions, participants predominantly produced gapped coordinations (Fig. 2), lending no support to H2.

Experiment 2 (acceptability rating, N = 49) tested the readers' sensitivity to the Verb Predictability (high/low) in C2 relative to the Form of the coordination (nonelliptical/Gapping). We used the 24 item sets with the largest effect size of Object Number from Exp. 1 (as indicated by *phi* coefficients) and 64 fillers. We expected that Gapping is rated as more acceptable compared to nonelliptical coordinations when the C2 verb is more predictable (Form \times Verb Predictability interaction). The data (analyzed with CLMMs, Christensen, 2022) showed no Verb Predictability \times Form interaction ($z = -0.30, p > 0.05$) (Fig. 3), i.e. our hypothesis was not borne out. There was a main Verb Predictability effect ($z = 9.01, p < 0.001$): Ratings were higher with more predictable C2 verbs.

Experiment 3 (self-paced reading, N = 96) investigated how the predictability of the C2 verb is related to the processing of Gapping. The same set of items (extended by a causative clause as spillover) and fillers as in Exp. 2 were presented word by word in a moving window paradigm. We expected that Gapping is processed easier with more predictable C2 verbs, i.e. faster RTs immediately after the ellipsis site (Verb Predictability effect for Gapping). An LMM analysis with nested contrasts revealed no Verb Predictability effect for Gapping ($t = -0.53, p > 0.05$) (Fig. 4), contrary to our prediction. There was a significant Verb Predictability effect for nonelliptical coordinations ($t = 2.04, p < 0.05$): The C2 object noun was read slower the more predictable the C2 verb was.

Discussion. We found no evidence that Gapping is used to avoid predictable material. The strong production preference for Gapping (Exp. 1) and the processing disadvantage for nonelliptical conjuncts with more predictable verbs (Exp. 3) suggest that Gapping might be the expected form for coordinations that license it. Thus, with predictable C2 verbs, the nonelliptical form becomes highly

¹All models included the maximal random effects structure. If the model did not converge, the RE structure was reduced by excluding the effects with the lowest variance (model comparisons with likelihood ratio tests). In all models, Verb Predictability was coded as continuous variable, using the cloze probabilities gathered in Exp. 1.

unpredictable and harder to process (Exp. 3). The lack of difference between the forms in Exp. 2 might be due to subjects focusing on the plausibility in context when rating the coordinations.

- (1) Anna is knitting a sweater, and Max **(is knitting)** a scarf.
- (2) *Die Anna und der Max haben im Bastelladen (Wolle | Wolle und the.NOM Anna and the.NOM Max have in.the craft.store wool | wool and Origamipapier) gekauft.*
origami.paper bought.
*Die Anna strickt einen Pulli und der Max **(strickt)** einen Schal.*
the.NOM Anna knits a.ACC sweater and the.NOM Max knits a.ACC scarf

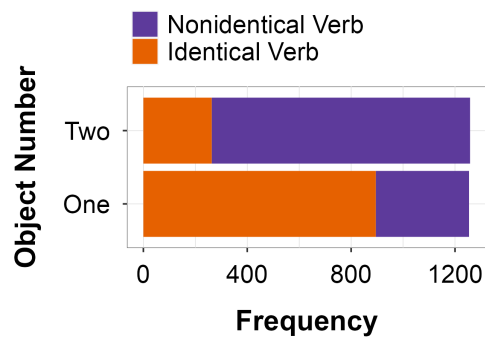


Fig. 1. Frequency of identical continuations

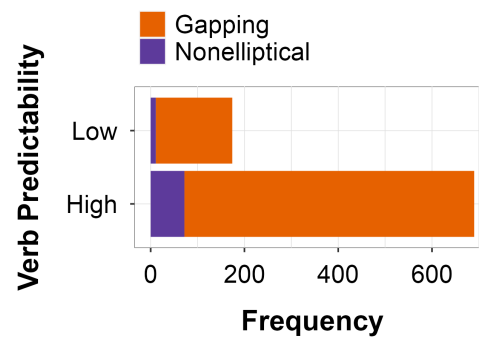


Fig. 2. Gapping ratio in the identical continuations

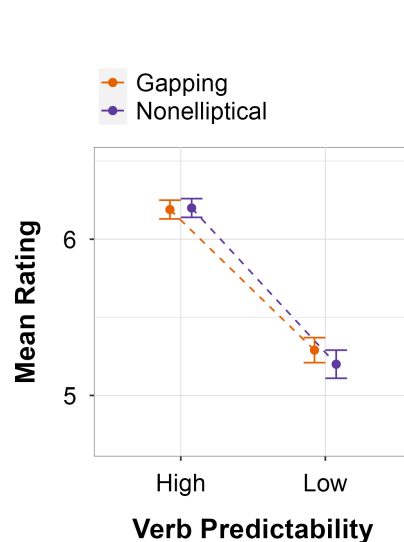


Fig. 3. Mean ratings per condition

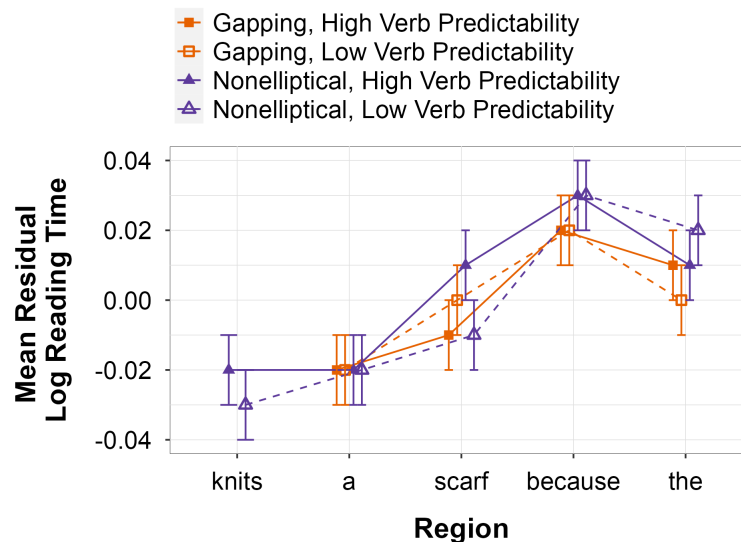


Fig. 4. Mean residual log reading times per region and condition

References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Christensen, R. H. B. (2022). Ordinal—Regression Models for Ordinal Data [R package version 2023.12-4]. <https://CRAN.R-project.org/package=ordinal>
- Fenk, A., & Fenk, G. (1980). Konstanz Im Kurzzeitgedächtnis-Konstanz Im Sprachlichen Informationsfluß. *Zeitschrift für experimentelle und angewandte Psychologie*, 27(3), 400–414.
- Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. C. Platt, & T. Hofmann (Eds.), *Advances in Neural Information Processing Systems* (pp. 849–856, Vol. 19). MIT Press.
- Ross, J. R. (1967). *Constraints on variables in syntax* [phd]. Massachusetts Institute of Technology.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.