Verb phrase ellipsis avoids troughs in the ID profile: An information-theoretic account to VPE based on evidence from rating and reading time data

Lisa Schäfer, Robin Lemke, Heiner Drenhaus and Ingo Reich (Saarland University)
lisa.schaefer@uni-saarland.de

Background. While the grammar of verb phrase ellipsis (VPE) is well-studied from a theoretical and a psycholinguistic perspective (see Reich (2011); Phillips and Parker (2014) for an overview), this is not true for the question of why a speaker uses an ellipsis at all. However, when contrasting the full form (1a) with the corresponding VPE (1b) an obvious hypothesis is that the speaker follows a pragmatic imperative for rational communication: Avoid redundancy!

(1) a. John played football in the backyard and Bill played football in the backyard, too.
b. John played football in the backyard and Bill did, too.

Account. This strategy may be modeled by the information-theoretic concept of Uniform Information Density (UID) (Levy and Jaeger, 2007). According to UID, speakers tend towards distributing information as uniformly as possible across utterances avoiding troughs in the information density (ID) profile (Fig. 1), i.e. regions with continuously low information. Troughs are caused by redundant information like the second conjunct of (1a) and are the longer the longer the redundant part is. We conducted an acceptability rating and a self-paced reading task that allow us to correlate preferences for VPE with cognitive effort indexed by reading times.

Exp. 1. We conducted the rating study in a 2 x 2 (FORM: ellipsis vs. full form x LENGTH: short (no PP) vs. long (with PP)) within-subjects design. In the short condition we presented a plain VP (2a & 2c), that was expanded by a PP in the long condition (2b & 2d). We expect a relatively higher preference for VPE in the long condition due to the longer trough in the ID profile. 41 British English native speakers recruited via Prolific Academic (Palan and Schitter, 2018) rated 32 items like (2) and 72 fillers (i.a. coordinated structures with two distinct VPs (3)) on a 7-point Likert scale (7 = completely natural). We analyzed our data with CLMMs (`ordinal`, Christensen (2018)) in `R` with a full random effects structure (Barr et al., 2013) and compared models with likelihood ratio tests: A significant main effect of LENGTH (χ^2 = 27.92, p < .001) shows that overall items with longer VPs were rated worse (Fig. 2). A significant FORM:LENGTH interaction (χ^2 = 7.66, p < .01) indicates that the full form was rated particularly worse in the long condition. This is in line with our hypothesis: Speakers prefer VPE the more the more redundant the repetition of the antecedent VP would be.

Exp. 2. Our account predicts that this preference is caused by the intention to avoid troughs in the ID profile. We used a self-paced reading study on the full forms (2a vs. 2b) to test whether the redundant VP indeed creates such a trough in the second conjunct. We expected this trough to be larger in the more redundant long condition, which would be indicated by a larger difference in average reading time between the first and the second conjunct in the long condition. Using IBEX (Drummond, 2016), the items and fillers from exp. 1, extended by a spillover region, were presented word-by-word and centered to 87 native speakers of British English recruited via Prolific Academic. Our dependent variable were the cumulated reading times per CONJUNCT (italicized in (2)) which we residualized per subject (Gibson and Levy, 2016). We analyzed the data with linear mixed effect models (`lme4`, Bates et al. (2015)) in `R` with random intercepts for subjects and items and compared models with likelihood ratio tests. A significant main effect of CONJUNCT (χ^2 = 159.18, p < .001) (Fig. 3) shows that the second conjunct was generally read faster than the first. A significant interaction between LENGTH:CONJUNCT (χ^2 = 63.04, p < .001) indicates that the second conjunct was especially faster in the long condition. This supports our hypothesis: The longer VP is more redundant and, thus, creates a longer trough, which is reflected in faster reading times.

Discussion. Our data provide further evidence for the effect of UID on encoding preferences: Speakers omit redundant parts of the utterance and use ellipsis to avoid troughs in the ID profile. This preference for omission increases the larger the redundant part is.

(2) a. John *played football* and Bill *played football*, too. (full form, short (no PP))

b. John *played football in the backyard of the house* and Bill *played football in the backyard of the house*, too. (full form, long (with PP))

c. John played football and Bill did, too. (ellipsis, short (no PP))

d. John played football in the backyard of the house and Bill did, too. (ellipsis, long (with PP))

(3) Ann packed provisions on the night before the trip and Tom loaded the car on the night before the trip (since they wanted to depart early in the morning). (filler (with spillover))

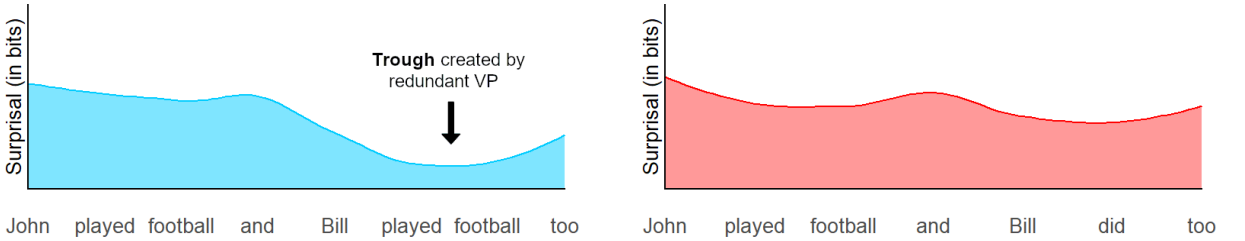

Figure 1: Hypothetical ID profiles for full form and VPE. While in the full form the repeated VP "played football" creates a trough this is not the case for the VPE in the elliptical counterpart.

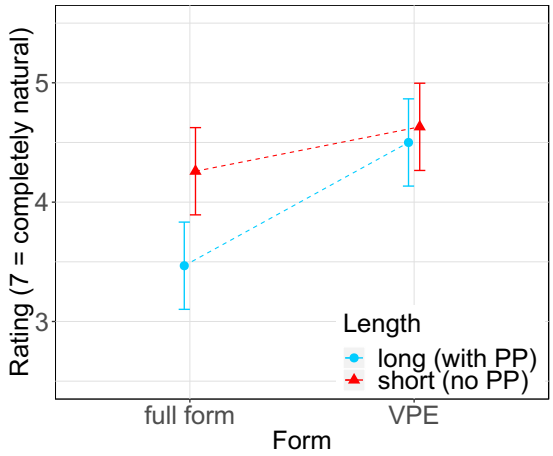

Figure 2: Mean ratings and 95% CIs for Exp. 1: VPEs are preferred, especially in the long condition.

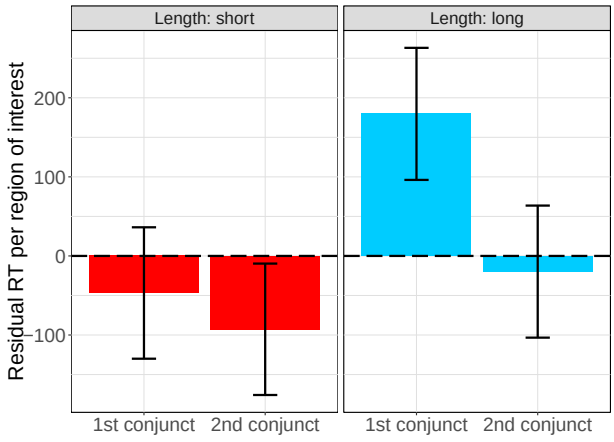

Figure 3: Cumulated residual RT per ROI and 95 % CIs for Exp. 2: 2nd conjuncts are read faster than 1st conjuncts, specially in the long condition.

References

Barr, D. J., Levy, R., Scheepers, C., and J. Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68:255–278.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Christensen, R. H. B. (2018). ordinal – Regression models for ordinal data. R package version 2018.8-25. http://www.cran.r-project.org/package=ordinal/.

Drummond, A. (2016). Ibex farm. Available: http://spellout.net/ibexfarm/. Accessed 2018 December 9.

Gibson, E. and Levy, R. (2016). An attempted replication of Hackl, Koster-Hale, Varvoutis (2012). Cornell ArXiv: arXiv:1605.00178 [q-bio.NC]; R-code and data available at https://osf.io/ t6anw.

Levy, R. P. and Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In Schölkopf, B., Platt, J. C., and Hoffman, T., editors, *Advances in Neural Information Processing Systems 19*, pages 849–856. MIT Press.

Palan, S. and Schitter, C. (2018). Prolific.ac – A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27.

Phillips, C. and Parker, D. (2014). The psycholinguistics of ellipsis. *Lingua*, 151:78–95. published online Nov 27, 2013.

Reich, I. (2011). Ellipsis. In Maienborn, C., Heusinger, K. v., and Portner, P., editors, *Semantics: An International Handbook of Natural Language Meaning.*, pages 1849–1874. de Gruyter, Berlin, New York.