

Can identity conditions on ellipsis be explained by processing principles?

Robin Lemke, Lisa Schäfer & Ingo Reich (Saarland University)

robin.lemke@uni-saarland.de

Background (Un)acceptable structural mismatches between antecedent and target are often used as evidence for semantic and/or syntactic identity conditions on ellipsis. For instance, Chung (2006: 83) argues that the acceptability of preposition omission under sluicing (1a), but not sprouting (1b), evidences a syntactic identity condition: All words that are omitted in the target (*John, danced, with*) must be present in the numeration of the antecedent.

- (1) a. John danced with somebody, but I don't know (with) who(m).
b. John danced, but I don't know *(with) who(m).

Proposal We hypothesize that at least some of the evidence for such specific licensing conditions can be explained by independently motivated processing mechanisms: Words are harder to process the less predictable they are (Hale, 2001; Levy, 2008) and speakers tend towards distributing processing effort uniformly across the utterance (Levy and Jaeger, 2007). This predicts mismatches to be more degraded the less predictable they are: In sluicing (1a), *somebody* provides an antecedent for the *wh* phrase, which has to be inferred in sprouting (1b) and therefore requires more processing effort. Furthermore, PP sluices (*with who*) will be easier to process in case of mismatches and sprouting, because the preposition *with* increases the likelihood of *who* and consequently spreads the information about the structural mismatch across more words.

Method We first ran an acceptability rating study in order to verify the pattern in (1) and then investigated processing difficulties on degraded sluices with a self-paced reading experiment. Both studies were web-based (IBEX) and conducted in German in a $2 \times 2 \times 2$ design: Construction (sluicing/sprouting), form of the Sluice (DP/PP) and Match, i.e. parallelism between sluice and antecedent (2). We tested only conditions (2a-f), because in the sprouting conditions (g,h), *Hans* would appear only in the 2nd conjunct's VP and cannot be recovered at all under ellipsis.

Rating study 48 subjects rated 24 items and 60 fillers on a 7-point Likert scale (7 = fully acceptable) in a web-based experiment. The analysis of the data (fig. 1) with CLMMs (Christensen, 2015) in R shows that mismatches under ellipsis are rated worse than in full forms both for sluicing and sprouting (Form:Ellipsis interactions, $\chi^2_{\text{sluice}} = 247.01$, $p < .001$, $\chi^2_{\text{sprout}} = 112.68$, $p < .001$), that mismatching DP sluices are worse than mismatching PPs (Form:Sluice, $\chi^2 = 6.16$, $p < .05$), and that mismatches under sprouting are worse than under sluicing ((2b) v. (2f), $\chi^2 = 27.84$, $p < .001$). This pattern is in line both with Chung's and our account.

Self-paced reading study 48 subjects read the full forms, i.e., including the material in brackets in (2a-f) and 60 fillers in a masked self-paced reading paradigm. We analyzed the residualized reading times (fig. 2) of the *wh* phrase and VP (e.g. (*mit*) *Hans getanzt hat*), which we used as spillover region, with linear mixed effects models (lme4, Bates et al. (2015)) in R. For both DP and PP sluices we find that matching sluices were read significantly faster than mismatching sluices ($\chi^2_{\text{DP}} = 3.98$, $p < .05$, $\chi^2_{\text{PP}} = 14.69$, $p < .001$). However, unlike the rating data might suggest, DP mismatches were not read more slowly under sprouting than under sluicing ($\chi^2 = 1.93$, $p < .1$).

Discussion Our experiments show that both DP and PP mismatches are degraded and harder to process than matching sluices. This is expected under our processing-based, but not under Chung's account. The acceptability penalty for sprouting mismatches as compared to sluicing is not reflected in significantly longer reading times. We are currently addressing this with cloze studies investigating whether the main verb (*to dance* in (1)) might serve as a cue towards a matching second conjunct in sprouting and thus facilitate the processing of mismatching second conjuncts.

- (2)
- a. Hans hat mit jemandem getanzt, aber ich weiß nicht, mit wem (Hans getanzt hat). PP, sluice, match
Hans has with someone danced but I know not with who_{dat} (Hans danced has).
'Hans danced with someone, but I don't know, with who(m) (Hans danced).'
 - b. Hans hat mit jemandem getanzt, aber ich weiß nicht, wer (mit Hans getanzt hat). DP, sluice, mism.
Hans has with someone danced, but I know not who_{nom} (with Hans danced has).
'Hans danced with someone, but I don't know, who (danced with Hans).'
 - c. Jemand hat mit Hans getanzt, aber ich weiß nicht, wer (mit Hans getanzt hat). DP, sluice, match
someone has with Hans danced but I know not who_{nom} (with Hans danced has).
'Someone danced with Hans, but I don't know, with who(m) (danced with Hans).'
 - d. Jemand hat mit Hans getanzt, aber ich weiß nicht, mit wem (Hans getanzt hat). PP, sluice, mism.
someone has with Hans danced but I know not with who_{dat} (Hans danced has).
'Someone danced with Hans, but I don't know, who (danced with Hans).'
 - e. Hans hat getanzt, aber ich weiß nicht, mit wem (Hans getanzt hat). PP, sprout, match
Hans has danced but I know not with who_{dat} (Hans danced has).
'Hans danced, but I don't know, with who(m) (Hans danced).'
 - f. Hans hat getanzt, aber ich weiß nicht, wer (mit Hans getanzt hat). DP, sprout, mism.
Hans has danced, but I know not who_{nom} (with Hans danced has).
'Hans danced, but I don't know, who (danced with Hans).'
 - g. Jemand hat getanzt, aber ich weiß nicht, wer (mit Hans getanzt hat). DP, sprout, match
someone has danced but I know not who_{nom} (with Hans danced has).
'Someone danced, but I don't know, with who(m) (danced with Hans).'
 - h. Jemand hat getanzt, aber ich weiß nicht, mit wem (Hans getanzt hat). PP, sprout, mism.
someone has danced but I know not with who_{dat} (Hans danced has).
'Someone danced, but I don't know, who (danced with Hans).'

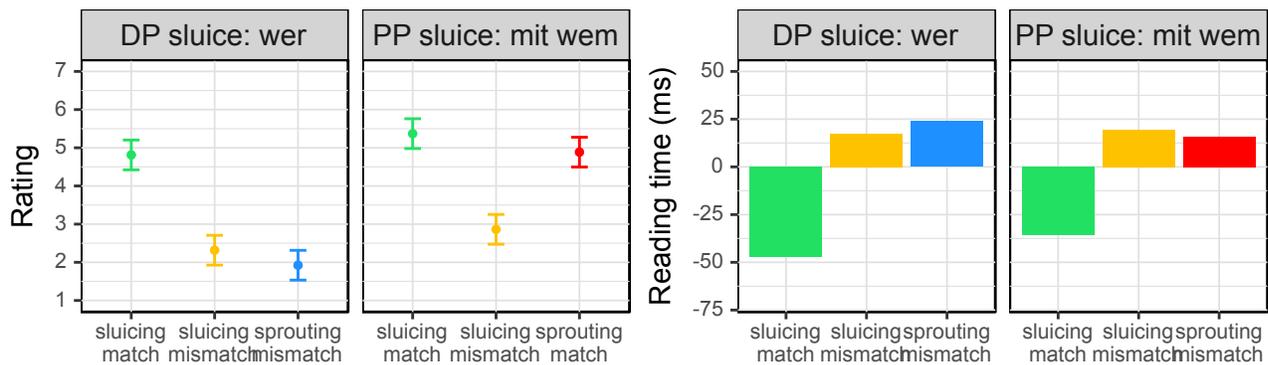


Figure 1 Mean ratings + 95% confidence intervals for the ratings on the elliptical utterances in the rating study. Figure 2 Mean residualized reading times for the corresponding full forms in the self-paced reading study.

References 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. (2015). Ordinal – Regression Models for Ordinal Data. R package version 2015.6-28. <http://cran.r-project.org/package=ordinal/>. • Chung, S. (2006). Sluicing and the Lexicon: The Point of No Return. *Annual Meeting of the Berkeley Linguistics Society*, 31(1). • Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL (Vol. 2)*, pages 159–166. • Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177. • Levy, R. P. and Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In Schlökopf, B., Platt, J., and Hoffman, T., editors, *Advances in Neural Information Processing Systems*, pages 849–856. MIT Press.