

Testing a noisy channel account of ellipsis processing

Robin Lemke (robin.lemke@uni-saarland.de)¹

¹Department of Modern German Linguistics, Saarland University

Background In the syntactic literature, the (un)acceptability of ellipses is often being used as a diagnostic for their grammaticality. However, this reasoning has been challenged by [1, 3], who propose a noisy channel account, according to which some relatively acceptable ellipses are instances of acceptable ungrammaticality: Listeners interpret them using a repair mechanism required to process input corrupted by noise. Speakers who exploit this produce elliptical utterances to communicate more efficiently. While this reasoning seems plausible, it has not been shown that processing ellipsis does rely on such a repair mechanism. The current study tests this explicitly by modulating the error rate in the fillers. Since listeners anticipate deletion errors when processing noisy input [4], a higher error rate should make ungrammatical ellipses easier to process and more acceptable.

Experiments Like [3], I investigate this at preposition omission under sluicing in German by comparing **three ellipsis conditions** (1). While sluicing a PP (*mit wem*) should be grammatical, syntactic approaches like [2] predict prepositional case-marked NP sluices (*wem*) to be ungrammatical, while nominative sluices (*wer*) are expected to be more acceptable because they can be derived without preposition omission from a source like *who it was*. These conditions are presented across **three presentation conditions** (between subjects): (i) a baseline (no errors), (ii) 50% of the fillers containing ungrammatical function word deletions, (iii) a speeded version of (ii), where the stimuli were presented word-by-word with a fixed presentation time of 400ms/word (more incremental paradigm). The noisy channel account predicts that ungrammatical/marked NP sluices improve with a higher error rate, in dative prepositional case NPs, which requires less string edits than the nominative (1 deletion vs. 2 deletions and 1 insertion). In each of the three studies conducted online on PClbex, 48 subjects (Prolific) rated 24 items and 72 fillers in pseudorandomized order on a 7pt Likert scale (7= fully natural). In the non-speeded versions, the items were displayed as a whole.

Results The data (See Fig. 1) were analyzed with Cumulative Link Mixed Models in R. The two ternary variables were forward-coded to distinguish between effects of P-OMISSION and CASE-MARKING as well as between effects of ERRORS and SPEEDED presentation (See Tab. 1). The model contained fixed effects for all predictors and their interactions, the scaled and centered POSITION of the trial in the experiment, by-subjects and by-item random intercepts and by-item random slopes for P-OMISSION and CASE-MARKING. The analysis shows that P-omission is degraded ($z=19.92$, $p<0.001$) and that prepositional case-marking is preferred over nominative ($z=5.79$, $p<0.001$). Two-way interactions indicate that for the noise conditions that (i) the preference for PPs is stronger, while (ii) the preference for prepositional over nominative case is reduced (possible floor effect).

Interpretation The data do not confirm the prediction of the noisy channel account that a higher deletion rate improves ungrammatical ellipsis. Subjects reject ungrammatical ellipsis across all noise conditions and reject P-omission even more strongly in the noise conditions. Interestingly, the speeded presentation increases the acceptability of ungrammatical fillers (see Fig. 1), but not to the same extent the acceptability of P-omission. This might indicate that subjects are adapting to noise, but that P-omission is different from the unsystematic deletions of function words in the fillers. This calls for a different processing-based or syntactic approach to P-omission under sluicing.

- (1) Kevin hat mit jemandem telefoniert, aber ich weiß nicht, (mit wem / wem / Kevin has with somebody talked.on.phone but I know not with who.DAT who.DAT wer).
 who.NOM
 ‘Kevin has talked to somebody on the phone, but I don’t know (with) who(m)’.



Figure 1: Mean rating + 95% confidence intervals across experiments and conditions as well as fillers (with)out ungrammatical deletions

Level	P-OMISSION	CASE-MARKING	Level	ERRORS	SPEEDED
PP	2/3	1/3	Baseline	-2/3	-1/3
NP, DAT	-1/3	1/3	50% errors	1/3	-1/3
NP, NOM	-1/3	-2/3	Errors, speeded	1/3	2/3

Table 1: Forward coding of the two ternary independent variables

	Estimate	Std. Error	z-value	Pr(> z)	
Errors	0.07	0.26	0.26	>0.7	
Speeded	0.24	0.26	0.90	>0.3	
P-omission	3.72	0.19	19.92	<0.001	***
Case-marking	0.44	0.08	5.79	<0.001	***
Position	-0.17	0.03	-5.06	<0.001	***
Errors:P-omission	0.91	0.21	4.52	<0.001	***
Speeded:P-omission	0.15	0.21	0.72	>0.4	
Errors:Case-marking	-0.45	0.19	-2.41	<0.05	*
Speeded:Case-marking	0.34	0.19	1.84	0.06	.

Table 2: Fixed effects in the CLMM predicting acceptability

References

- [1] Leon Bergen and Noah Goodman. The strategic use of noise in pragmatic reasoning. *Topics in Cognitive Science*, 7(2):336–350, 2015. doi: 10.1111/tops.12144.
- [2] Jason Merchant. *The Syntax of Silence. Sluicing, Islands, and the Theory of Ellipsis*. Oxford University Press, New York, 2001.
- [3] Emilia Molimpakis. *Accepting Preposition-Stranding under Sluicing Cross-Linguistically; a Noisy-Channel Approach*. PhD thesis, UCL, 2019.
- [4] Rachel Ryskin, Richard Futrell, Swathi Kiran, and Edward Gibson. Comprehenders model the nature of noise in the environment. *Cognition*, 181:141–150, December 2018. doi: 10.1016/j.cognition.2018.08.018.