Delete me if you can. Predictable words are difficult to process if they could have been omitted

| Bozhidara Hristova, Robin Lemke, Lisa Schäfer, Heiner Drenhaus, Ingo Reich (Saarland University) | bozhidara.hristova@uni-saarland.de

It is well attested that comprehenders' expectancy for a word is a reliable predictor for its processing ease, with more predictable expressions being processed more easily (Demberg & Keller, 2008). At the same time, speakers tend to omit predictable words in order to communicate more efficiently as long as this is grammatically licensed, which makes predictable content more likely to be omitted (Levy & Jaeger, 2007). Since comprehenders anticipate omissions (Yoshida et al., 2013), and the omission of highly predictable words reduces the likelihood of actually encountering them, their unexpected overt realization should be associated with increased processing effort. There is evidence that utterances expressing highly predictable propositions are harder to process (Rohde et al., 2021). However, there is a lack of studies investigating the effect of omission expectations on syntactic reduction phenomena. Are predictable words easier to process even if they can be elided, or do they lead to processing difficulties because comprehenders expect an omission?

Experimental rationale We present two self-paced reading (SPR) studies indicating that comprehenders process more predictable words with less ease if there is a reduced option available. We focus on the finite verb in the second conjunct (C2) of parallel coordinations in German (1), i.e. coordinations with the same verb in both conjuncts. The C2 verb of parallel coordinations can be either realized (nonelliptical structure) (1a) or omitted (gapping) (1b). If omissions are more strongly expected with more predictable content, we hypothesized that nonelliptical coordinations (1a) will be harder to process if the C2 verb (e.g., *strickt* 'is knitting' in (2)) is more predictable, i.e. slower RTs at the C2 verb and its spillover. We investigated this hypothesis with two different predictability manipulations, varying the predictability of a word through i) context; and ii) memory decay. For both manipulations, we conducted an online written production study to ensure the validity of the predictability manipulation and to norm the stimuli. For the SPR studies, we used the word-by-word moving-window paradigm.

Predictability manipulation through context (Exp. 1, SPR, n = 96) We modulated the predictability of the verb through the number of objects in context (OBJECT NUMBER, one vs two) (2). Mentioning two objects rather than one decreases the probability of the C2 verb (on average, from 0.76 to 0.20) since readers expect the second object to be related to another action. In a 2 × 2 design, we crossed OBJECT NUMBER with the form of the coordination (FORM, nonelliptical vs gapping). An analysis with linear mixed models (LMM, Bates et al. 2015) and nested contrasts (OBJECT NUMBER nested within FORM) revealed a significant effect of OBJECT NUMBER at the object noun of the nonelliptical structures (e.g., *Schal* 'scarf' in (2)): The C2 object noun was read slower in the one-object (predictable) condition (t = 2.00, p < 0.05). There was no effect of OBJECT NUMBER in the gapped coordinations (t = -1.28, p > 0.2).

Predictability manipulation through memory decay (Exp. 2, SPR, n = 96) The predictability of a word is dependent on the memory representations of the context (Futrell et al. 2020). Since the predictability of the C2 verb is based on the same verb being used in the first conjunct, the C2 verb becomes less predictable as the distance between the two conjuncts increases and the memory representations of the first conjunct become less accurate. Therefore, we manipulated the predictability of the C2 finite verb by varying the distance between the two conjuncts through a parenthetical clause (DISTANCE) (3). In a 2 × 2 design, DISTANCE (short vs long) was crossed with FORM (nonelliptical vs gapping). An LMM analysis with nested contrasts (DISTANCE nested within FORM) showed a significant effect of DISTANCE in the nonelliptical items, so that the C2 object noun was read slower in the short-distance (predictable) condition than in the long-distance condition (t = 2.20, p < 0.05). There was no effect of DISTANCE in the gapping items (t = 0.61, p > 0.5).

Discussion Our results suggest that predictable words are harder to process if a reduced form is available. We assume that this effect stems from the comprehenders' expectations about the form of expressions: The realization of a word is less likely if the word is more predictable. To our knowledge, our studies are the first to show that the availability of a reduced alternative influences the effect of predictability on sentence processing. The results indicate that the presence of form alternatives should be considered in models of comprehension.

(1) a. *Mia is knitting a sweater, and Jan* ***is knitting*** *a scarf.* (nonelliptical form)
b. *Mia is knitting a sweater, and Jan a scarf.* (gapping)

(2) a. **One** object in context (**predictable** C2 verb):

Die	*Mia*	*und*	*der*	*Jan*	*haben*	*im*	*Bastelladen*	***Wolle***	*gekauft.*				
the	Mia	and	the	Jan	have	in	craft.store	wool	bought				
Die	*Mia*	*strickt*	*einen*	*Pulli*	*und*	*der*	*Jan*	***⟨strickt⟩***	*einen*	*Schal,*	weil	der	Winter
the	Mia	knits	a.ACC	sweater	and	the.NOM	Jan	knits	a	scarf	because	the	winter
kalt	*sein*	*wird.*											
cold	be	will											

b. **Two** objects in context (**unpredictable** C2 verb):

Die	*Mia*	*und*	*der*	*Jan*	*haben*	*im*	*Bastelladen*	***Wolle***	***und***	***Origamipapier***	*gekauft.*		
the	Mia	and	the	Jan	have	in	craft.store	wool	and	origami.paper	bought		
Die	*Mia*	*strickt*	*einen*	*Pulli*	*und*	*der*	*Jan*	***⟨strickt⟩***	*einen*	*Schal,*	weil	der	Winter
the	Mia	knits	a.ACC	sweater	and	the.NOM	Jan	knits	a	scarf	because	the	winter
kalt	*sein*	*wird.*											
cold	be	will											

(3) a. **Short** distance to antecedent (**predictable** C2 verb):

Diesen	*Herbst,*	***soweit***	***ich***	***das***	***am***	***Telefon***	***richtig***	***verstanden***	***habe,***	*strickt*	*die*	*Mia*	
this	autumn	as.far.as	I	that	on.the	phone	correctly	understood	have	knits	the	Mia	
einen	*Pulli*	*und*	*der*	*Jan*	***⟨strickt⟩***	*einen*	*Schal,*	*weil*	*der*	*Winter*	*kalt*	*sein*	*wird.*
a.ACC	sweater	and	the.NOM	Jan	knits	a	scarf	because	the	winter	cold	will	be

b. **Long** distance to antecedent (**unpredictable** C2 verb):

Diesen	*Herbst*	*strickt*	*die*	*Mia*	*einen*	*Pulli,*	***soweit***	***ich***	***das***	***am***	***Telefon***	***richtig***	
this	autumn	knits	the	Mia	a.ACC	sweater	as.far.as	I	that	on.the	phone	correctly	
verstanden	***habe,***	*und*	*der*	*Jan*	***⟨strickt⟩***	*einen*	*Schal,*	*weil*	*der*	*Winter*	*kalt*	*sein*	*wird.*
understood	have	and	the.NOM	Jan	knits	a	scarf	because	the	winter	cold	be	will

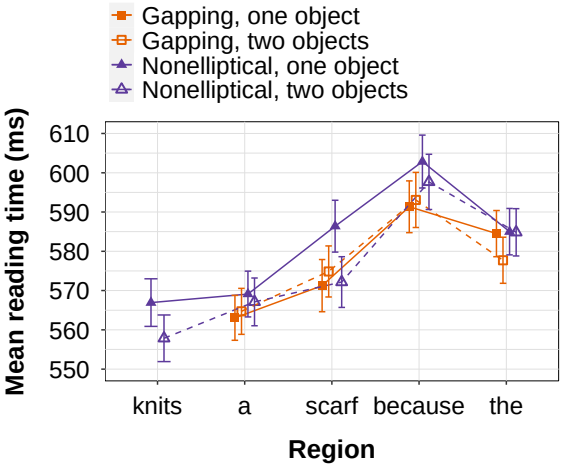

Fig. 1 Mean reading times and standard errors per region and condition for Exp. 1

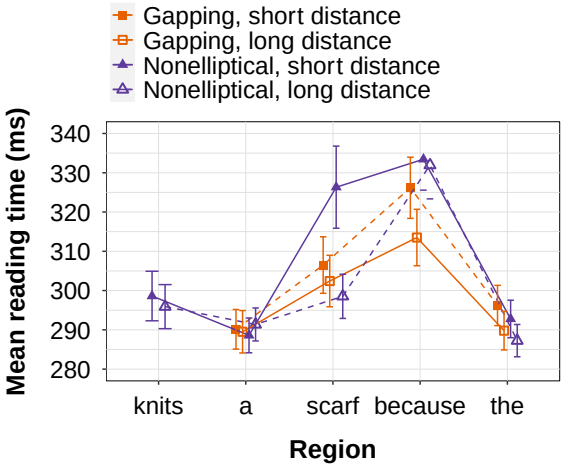

Fig. 2 Mean reading times and standard errors per region and condition for Exp. 2

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.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210.

Futrell, R., Gibson, E., & Levy, R. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science, 44*.

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.

Rohde, H., Futrell, R., & Lucas, C. G. (2021). What's new? A comprehension bias in favor of informativity. *Cognition*, *209*.

Yoshida, M., Dickey, M. W., & Sturt, P. (2013). Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. *Language and Cognitive Processes*, *28*(3), 272–302.