A pre-registered large-sample investigation of similarity-based interference in English, German and Russian

Daniela Mertzen[1], Anna Laurinavichyute[1], Brian Dillon[2] & Shravan Vasishth[1]
[1]University of Potsdam, Germany; [2]University of Massachusetts Amherst, MA, USA
mertzen@uni-potsdam.de

Cue-based parsing theories [1-3] assume that dependency formation during real-time sentence processing relies on cue-based retrieval of syntactic encodings in working memory. This retrieval mechanism is prone to similarity-based interference which can occur when there are multiple items in memory that are similar to the target of retrieval. Van Dyke & McElree [4] showed that these cue-dependent interference effects can even stem from similar sentence-external lexical items that were encoded in working memory for a concurrent memory task.

[4] tested this in a dual task (recall + self-paced reading) experiment in English using a 2 (memory load) x 2 (interference) design. In load conditions, participants were asked to memorize three nouns, and recall them after reading the target sentence (see Table 1: a, b). The *memory nouns* are either plausible objects of the critical relative-clause verb, or they are not (e.g., in Table 1, memory nouns are plausible objects of (b) *fixed* but not (a) *sailed*). In no load conditions (c, d), nothing had to be memorized. An interference effect was observed in the form of slower reading times at the critical verb when memory nouns were plausible objects (e.g., (b) *fixed* was read slower than (a) *sailed*). In no load conditions, no reading time difference was observed. A subsequent replication found this interference effect only in offline comprehension data [5].

We re-examined this important finding with a larger-sample eye-tracking study in English (N=75) and examined the cross-linguistic generality of the effect by testing two languages (German, N=122; Russian, N=120) that use overt case marking; we reasoned that case marking might reduce interference effects in real-time sentence comprehension. To establish whether interference effects are conditional on processing depth, we additionally manipulated comprehension question complexity [6,7], inducing *deep* and *shallow* processing (see Table 2 for overview of experiments).

Our data were analyzed using maximal Bayesian linear mixed effects models with load, interference and their interaction as fixed effects. We pre-registered our prediction of an interaction (with a positive sign) in *total fixation time* (TFT). To evaluate the effect estimates, we also pre-registered a *null region* [8,9] of $\pm$ 20 ms around zero for TFT which counts as "no effect" (Fig. 1). This range is based on previous studies' effect magnitudes in eye-tracking measures (see [10], p. 162). If the empirical estimate lies outside the null region this would be interpreted as evidence for an effect, a partial overlap with the null region would suggest inconclusive evidence for an effect, and if the estimate falls entirely within the null region, this is interpreted as no evidence for an effect.

Fig. 1 shows means of the posterior distributions with their 95% credible intervals (CrI) for the Load $\times$ Interference interaction: we saw no indication of the predicted interaction in any of the three languages, and no effects were contingent on processing depth. All effect estimates fall within the pre-defined null region. Recall accuracies were relatively high across experiments (Table 3), demonstrating participants did not merely disregard the recall task.

Overall, results from our pre-registered analyses do not lend support to the hypothesis that sentence-external items in working memory can interfere with retrieval during sentence processing, in the tested languages. It is possible that the interference effect caused by sentence-external distractors is very small [11] and difficult to detect, or that interfering distractors play a role only when they appear within a sentence; the latter prediction is currently being tested.

References: **[1]** Lewis & Vasishth (2005). Cog Sci. **[2]** McElree (2000). J. Psycholinguist. Res. **[3]** Van Dyke & Lewis (2003). JML. **[4]** Van Dyke & McElree (2006). JML. **[5]** Van Dyke et al. (2014). Cognition. **[6]** Ferreira, Ferraro & Bailey (2002). Curr. Dir. Psychol. Sci. **[7]** Swets et al. (2008). Mem Cognition. **[8]** Freedman, Lowe & Macaskill (1984). Biometrics. **[9]** Hobbs & Carlin (2008). J. Biopharm. Stat. **[10]** Vasishth et al. (2018). JML. **[11]** Van Dyke & McElree (2011). JML.

Table 1: English example item (adapted for our study from Van Dyke & McElree, 2006)

Memory load conditions:
table *sink* *truck*

a. No interference
The boat that the guy who lived by the sea sailed in the morning was very old.

b. Interference
The boat that the guy who lived by the sea fixed in the morning was very old.

No memory load conditions:
— — —

c. No interference
The boat that the guy who lived by the sea sailed in the morning was very old.

d. Interference
The boat that the guy who lived by the sea fixed in the morning was very old.

Table 2: Experiments testing design shown in Table 1. Complex refers to the version of the study that attempted to induce 'deep processing' and simple refers to the version of the study inducing 'shallow' processing. The same participants saw both versions 7 to 21 days apart.

Study	Expt version	Subjects	Items
English	Complex	75	40
	Simple		40
German	Complex	122	40
	Simple		40
Russian	Complex	120	40
	Simple		40

Table 3: Recall accuracy of memory words for each language and version. Displayed is the respective accuracy in % according to a lenient criterion (either two or three words had to be correctly recalled in any order). Int = interference.

Study	(a) Load, No int	(b) Load, Int
English, complex	77	75
English, simple	86	84
German, complex	84	85
German, simple	85	86
Russian, complex	85	84
Russian, simple	91	92

Accuracy (%)

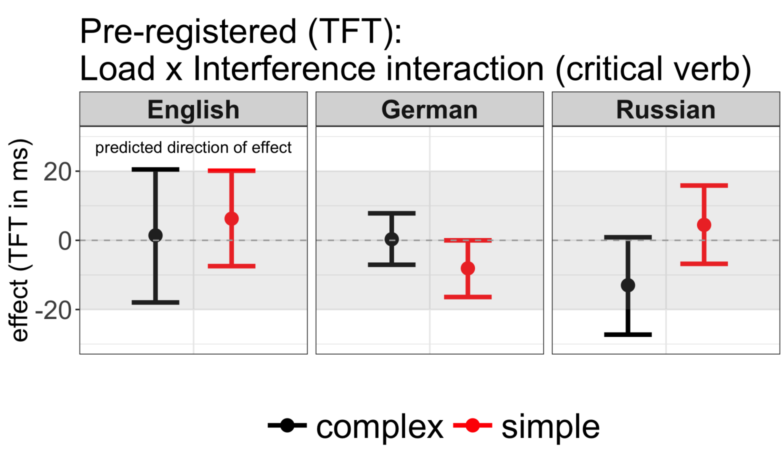

Figure 1: Posterior means with 95% credible intervals for the Load × Interference interaction (total fixation time, back-transformed from the log- to the ms-scale) for English, German and Russian. Complex refers to the version of the study that attempted to induce 'deep processing' using difficult comprehension questions and simple refers to the version of the study inducing 'shallow' processing. The interaction was predicted to have a positive sign. The null region is highlighted in gray.