Cue confusion and distractor prominence explain inconsistent effects of retrieval interference in human sentence processing

Felix Engelmann (felix.engelmann@uni-potsdam.de)
Department of Linguistics, University of Potsdam, Karl-Liebknecht-Str. 24/25
14476 Potsdam, Germany

Lena Jäger (lena.jaeger@uni-potsdam.de)
Department of Linguistics, University of Potsdam, Karl-Liebknecht-Str. 24/25
14476 Potsdam, Germany

Shravan Vasishth (vasishth@uni-potsdam.de)
Department of Linguistics, University of Potsdam, Karl-Liebknecht-Str. 24/25
14476 Potsdam, Germany

Keywords: sentence processing; retrieval; interference; ACT-R

The cognitive mechanisms underlying the processing of non-adjacent syntactic dependencies are critical for the understanding of human language processing. For instance, a verb needs to be syntactically and semantically integrated with its subject, or a reflexive like *himself* needs to be syntactically bound by its antecedent before it can be assigned any meaning. Thus, when processing the second part of a syntactic dependency, the parser needs to retrieve the corresponding first part of this dependency. The mechanisms underlying these syntactically triggered retrieval processes have drawn considerable attention in psycholinguistic research. Lewis and Vasishth (2005) (LV05) developed a model of sentence processing which is based on the general cognitive architecture ACT-R (Anderson et al., 2004). This model assumes a content-addressable memory in which cue-based retrieval processes are subject to similarity-based interference from (partially) cue-matching distractors. The LV05 model has widely been used to explain interference effects observed in the processing of syntactic dependencies such as reflexive-antecedent or subject-verb dependencies. Although the model is able to capture some of the empirically observed effects, there is a range of data the model is unable to explain. We propose to extend the LV05 model by two independently motivated assumptions, namely *cue confusion* and *activation-sensitive interference*. We demonstrate that this extended model explains a wide range of empirically observed effects the original LV05 model does not account for.

The LV05 model predicts that when retrieving the left part of a dependency (the target), a syntactically inaccessible noun phrase (distractor) that overlaps in features with the target noun phrase causes similarity-based interference, which leads to slowed processing (i.e., inhibitory interference). This is predicted, e.g., in the retrieval of a reflexive's antecedent as in (1) in Table 1, where the stereotypical gender on the target *surgeon* and on the distractor *Jonathan* both match the gender cue on the reflexive. By contrast, in (2), the stereotypical gender of the target *surgeon* mismatches the gender cue at the reflexive; here, a matching distractor is predicted to speed up processing by luring the parser into erroneous retrievals (facilitatory interference). Both effects are attested (e.g., Pearlmutter, Garnsey, & Bock, 1999; Badecker & Straub, 2002).

However, some studies have found facilitatory interference where inhibition was expected, and vice versa; other studies have failed to find interference effects. We developed a computational model extending LV05 by two independently motivated principles that can account for these apparently contradictory results. We show this in simulations that reproduce the patterns that were seen in a large-scale literature review.

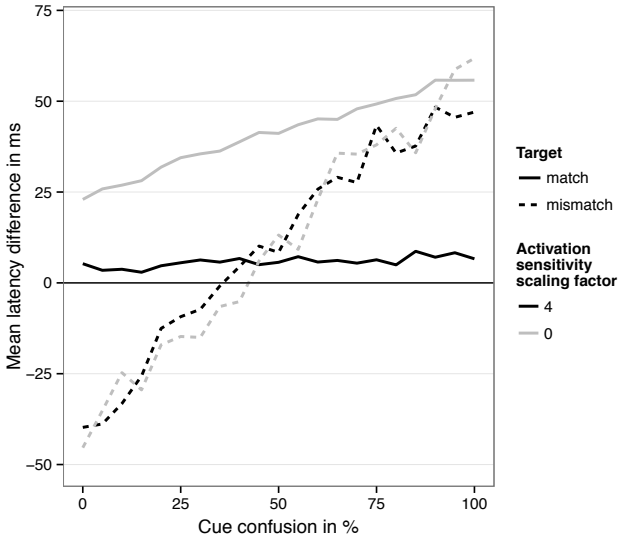

Figure 1: Predicted interference effects (interference condition - no-interference condition) by cue confusion for models without activation-sensitive interference (gray lines) and with activation-sensitivity scaling factor 4 (black lines). Solid lines represent the conditions where the target matches the semantic cue, mismatch conditions are represented by dashed lines. The LV05 model's prediction is shown by the gray lines at cue confusion of 0%.

Table 1: Gender-match/mismatch design commonly used in psycholinguistic experiments investigating interference effects in reflexives; example from Sturt (2003).

Match type	Example	Prediction (LV05)
(1) Target-Match	The **SURGEON**$^{+masc}_{+c\text{-}com}$ who treated [Jennifer$^{-masc}_{-c\text{-}com}$/**Jonathan**$^{+masc}_{-c\text{-}com}$] had pricked **HIMSELF**$^{masc}_{c\text{-}com}$	inhibition
(2) Target-Mismatch	The SURGEON$^{-fem}_{+c\text{-}com}$ who treated [Jonathan$^{-fem}_{-c\text{-}com}$/**Jennifer**$^{+fem}_{-c\text{-}com}$] had pricked **HERSELF**$^{fem}_{c\text{-}com}$	facilitation

Principle 1: Cue confusion

We assume that a retrieval cue can be associated with more than one feature. The strength of this association is represented on a continuous scale and is shaped by experience. If two retrieval cues co-occur frequently in a certain retrieval environment, each of the two cues becomes associated also with the feature matched by the other cue. E.g., the Mandarin reflexive *ziji* invariantly cues for the feature pair $\{^{anim}_{c\text{-}com}\}$. This co-occurrence leads to a certain *crossed association* between *c-com* and *anim*. The same would hold for the *c-com* and *plur* cues in reciprocals. By contrast, English reflexives vary in number and gender: $\{^{fem/masc,\ plur/sing}_{c\text{-}com}\}$, resulting in a stronger one-to-one association rather than a crossed association between *c-com*, number, and gender. With crossed cue-feature associations, similarity-based interference can arise between memory items that do not share the same features. This explains the inhibitory interference effects observed in Target-Mismatch in Mandarin reflexives (Jäger, Engelmann, & Vasishth, subm.) and Hindi reciprocals (Kush & Phillips, 2014).

Independently of cue co-occurrence, we suggest that the associative strength between cues and features is modulated by working memory capacity: A strong one-to-one association is assumed to involve cognitive effort, hence readers with lower working memory capacity experience more crossed associations, leading to inhibitory interference in Target-Mismatch, even in English reflexives, as has been observed by Cunnings and Felser (2013).

Principle 2: Activation-sensitive interference

The strength of similarity-based interference is assumed to be scaled by the activation difference between target and distractor. E.g., in Target-Match, the target activation is much higher than the distractor activation because the target is a perfect match to the retrieval cues, which reduces the interference effect induced by the distractor. Thus, the following three patterns can be explained by distractor activation (prominence): (i) the well-known "grammatical asymmetry" (Wagers, Lau, & Phillips, 2009): interference effects are found more reliably in Target-Mismatch than in Target-Match; (ii) inhibitory interference increases in Target-Match when the distractor is more active, e.g., when it is in a more prominent subject position (Badecker & Straub, 2002); and (iii) facilitatory interference in Target-Match (e.g., Cunnings & Felser, 2013) due to fast misretrievals masking the similarity-based interference when the distractor has an even higher activation than the target.

Conclusion

In summary, we show in a computational model how two independently motivated principles that extend LV05's cue-based retrieval theory provide a principled explanation of hitherto unexplained patterns in the literature on interference in dependency processing: *Cue confusion* accounts for unexplained inhibitory interference, and *activation-sensitive interference* explains the conditions under which interference effects disappear

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004, October). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–60.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 748–769.

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes*, *28*(1-2), 188–219.

Kush, D., & Phillips, C. (2014). Local anaphor licensing in an SOV language: Implications for retrieval strategies. *Frontiers in Psychology*, *5*(1252).

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*, 427–456.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*, 542–562.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*, 206–237.