Exploring memory and processing through a gold standard annotation of Dundee

Cory Shain[1], Marten van Schijndel[1], Edward Gibson[2] and William Schuler[1] [1]Ohio State U. [2]MIT
shain.3@osu.edu

Theories of sentence processing such as Dependency Locality Theory (DLT) [2] predict memory-related processing difficulty proportionate to the number of discourse referents intervening between a noun or finite verb and its backward-looking dependency. This prediction has held in experiments using constructed stimuli [6]. However, naturalistic studies of memory effects [1] using the Dundee eye-tracking corpus [3] have shown negative effects for integration of automatically estimated dependencies when applied broadly, and only weak positive effects (when corrected for multiple trials) when applied more narrowly. One possible explanation for this is that automatically-parsed dependencies might introduce bias due to the difficulty of parsing longer sentences, which disproportionately contain long dependencies.

The experiment described here re-evaluates DLT using hand-corrected syntactic annotations of the Dundee corpus in an HPSG-like representation [4]. We use this annotation to evaluate DLT as a predictor of first-pass durations over an independently-motivated baseline of memory-independent factors [5].[1] Since human perception generally operates on a logarithmic scale (i.e. Weber-Fechner law) and to maintain comparability to [1], we log-transform the DLT predictor. Results were not significant (see Table 1), showing that the negative effect found by [1] in their initial experimental setup[2] may have been due to automatic parser errors. However, the positive effect predicted by DLT was not observed.

	Effect (ms)	p
DLT (orig)	-0.375	0.643
BothMod	-2.22	0.005

Table 1: Results

We then evaluate three DLT variants designed to better account for broad-coverage phenomena: VerbMod, CoordMod, and BothMod. VerbMod assigns finite verbs a cost of 2 energy units (not 1) and non-finite verbs a cost of 1 (not 0), since finite verbs might be more difficult to integrate than non-finite because they contain more detail (e.g. tense) to instantiate in working memory (see e.g. [6]). In CoordMod, total cost for coordinates equals that of their heaviest conjunct, and preceding conjuncts are skipped in the calculation of integration costs for discourse referents under coordination, following the notion that each sub-referent of a conjunction is integrated into a conjoined set which is finally integrated at the end of the conjunction. BothMod applies both modifications.

To avoid excessive multiple trials correction, we first test these three (log-transformed) DLT variants on an exploratory set of the Dundee corpus (every 3rd sentence), then evaluate only the version that most improves model fit (BothMod) on the remaining held-out sentences. Contrary to the predictions of DLT, the effect is significant and negative (see Table 1). Thus the negative integration cost observed in previous naturalistic studies cannot simply be reduced to an artifact of automatic parsing.

References

[1] V. Demberg and F. Keller. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2), 2008.

[2] E. Gibson. The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, language, brain*, 2000.

[3] A. Kennedy, et al. The Dundee corpus. In *12th European conference on eye movement*, 2003.

[4] L. Nguyen, et al. Accurate unbounded dependency recovery using generalized categorial grammars. In *COLING 2012*, 2012.

[5] M. van Schijndel and W. Schuler. Hierarchic syntax improves reading time prediction. In *NAACL 2015*, 2015.

[6] T. Warren and E. Gibson. The influence of referential processing on sentence complexity. *Cognition*, 85, 2002.

[1] Sentence position, word length, length of preceding saccade, whether the preceding token was fixated, cumulative 5-gram probability, and total surprisal, with by-subject random slopes for each of these and random intercepts by word.

[2] Using exploratory data, we attempted partial replication of the positive integration cost found by [1] by removing all tokens with an integration cost of 0. The only effect was to further reduce the significance of the predictor.