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]

shain.3@osu.edu

[1]The Ohio State University, [2]MIT

Introduction

Predictions of the Dependency Locality Theory (DLT) [3] have held in experiments using constructed stimuli [9], but effects have been weak or negative when applied broadly in naturalistic studies (e.g. [1]). We use hand-corrected syntactic annotations of the Dundee eye-tracking corpus [4] to evaluate the possibility that this is due to errors in automatic dependency estimation.

Baseline model:
Sentence position, word length, length of preceding saccade, cumulative 5-gram probability, and total surprisal, with by-subject random slopes for each of these and random intercepts by word.

Exploratory data:
Every third sentence of Dundee.

Held-out data:
All Dundee sentences not in exploratory data.

Experiment 1 - DLT on Gold Dundee

We fit first-pass durations on held-out data using a baseline [7] (see above) and log-transformed DLT integration cost. Non-significant results (Table 2) show the negative effect found initially by [1] may have been due to automatic parser errors. Note that the predicted positive correlation with reading times is not observed, either.

Experiment 2: Broad-coverage variants of DLT

We then tested three broad-coverage modifications to DLT (right). BothMod most improved model fit on exploratory data (Table 1), so it was evaluated on the remainder of the corpus. Contrary to DLT predictions, the effect is significantly negative (Table 2).

Results - Exp. 1 & 2

	Effect (ms)	p
DLT (orig)	-1.314	0.158
CoordMod	-1.983	0.042
VerbMod	-2.593	0.010
BothMod	-3.324	0.002

Table 1: Results on exploratory data

	Effect (ms)	p
DLT (orig)	-0.333	0.652
BothMod	-2.177	0.006

Table 2: Results on held-out data

A Hand-Corrected Syntactic Annotation of Dundee

We hand-corrected syntactic parses of the entire Dundee eyetracking corpus [4], using a derivative of the Nguyen et al. generalized categorial grammar for English [5] (using -a and -b for unsatisfied preceding and succeeding arguments, and -g and -r for non-local filler-gap and relative pronoun dependencies). These annotations allow non-local dependenceis (-g/-r) to be learned by a PCFG parser, used in the hand-correction process and in the calculation of surprisal. This annotation allows us to test syntax-based theories of sentence processing with a substantially reduced risk of spurious results due to incorrectly-estimated dependencies.

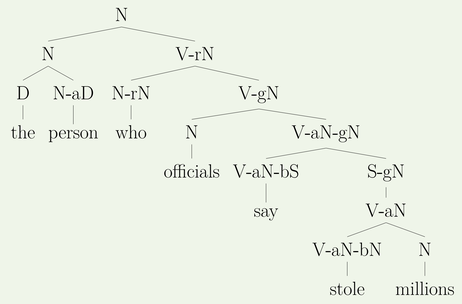

Figure 1: Nguyen et al. representation

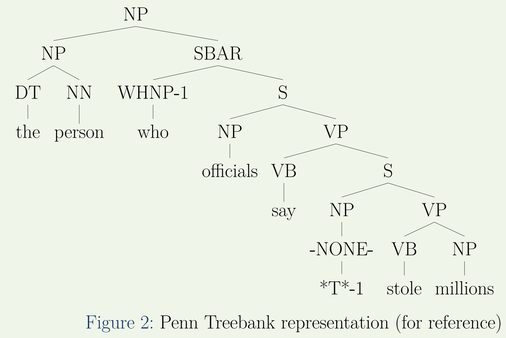

Figure 2: Penn Treebank representation (for reference)

To access the public release of this Dundee annotation, visit **http://go.osu.edu/golddundee**.

DLT Variants tested

We used the following variants of DLT (modifications designed to better account for broad coverage phenomena):

- **Variant 0: Unmodified DLT** Nouns and finite verbs incur an integration cost of 1 (for the word itself) plus 1 for each noun or verb intervening in its backward-looking dependency.

(1)

The	person	that	supervisors	and	co-workers	caught	stealing	millions	...
0	1	0	1	0	1	4	0	1	

- **Variant 1: VerbMod** Finite verbs introduce an integration cost of 2 rather than 1, and non-finite verbs introduce an integration cost of 1 rather than 0 [9].

(2)

The	person	that	supervisors	and	co-workers	caught	stealing	millions	...
0	1	0	1	0	1	6	1	1	

- **Variant 2: 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.

(3)

The	person	that	supervisors	and	co-workers	caught	stealing	millions	...
0	1	0	1	0	1	3	0	1	

- **Variant 3: BothMod** Both VerbMod and CoordMod are applied together.

(4)

The	person	that	supervisors	and	co-workers	caught	stealing	millions	...
0	1	0	1	0	1	5	1	1	

Conclusion

DLT is not a significant predictor of first pass reading times when evaluated over hand-corrected syntactic annotations in Dundee, suggesting that parser error may have played a role in earlier findings of significant negative effects based on automatic dependency estimation. However, an independently-motivated variant of DLT shows a significant facilitory effect, suggesting that the negative integration cost observed in previous naturalistic studies may not simply be an artifact of automatic parsing. This result is consistent with previous work that has found facilitation despite controlling for surprisal [8, 6].

Follow-Up Study 1: Sentence Intercept

To control for the possibility of sentence-level effects, we added a random intercept by subject:sentid pair to the baseline. The negative effect persists on exploratory data (see Table 3).

	Effect (ms)	p
DLT (orig)	-1.345	0.153
BothMod	-3.294	0.002

Table 3: Results with new baseline on exploratory data

Follow-Up Study 2: Amateur Novels Corpus

We ran follow-up study 1 on every 3rd sentence of the Amateur Novels Corpus [2] (fewer words but more subjects), again using hand-corrected syntactic annotations. Results are not significant for either DLT (original) or BothMod (see Table 4).

	Effect (ms)	p
DLT (orig)	0.095	0.963
BothMod	-0.366	0.869

Table 4: Results with new baseline on an exploratory set of the Amateur Novels Corpus

Perhaps the Dundee findings do not generalize well to a larger population because of the small sample size (10 subjects). However, since the Amateur Novels Corpus has fewer sentences than Dundee and those sentences are generally simpler, Amateur may lack either (i) statistical power or (ii) constructions which drive the Dundee result, or both. We leave this as an area for future research.

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] Stefan L. Frank, Irene Fernandez Monsalve, Robin L. Thompson, and Gabriella Vigliocco. Reading time data for evaluating broad-coverage models of English sentence processing. *Behavior Research Methods*, 45:1182–1190, 2013.

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

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

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

[6] Bruno Nicenboim, Shravan Vasishth, Carolina Gattei, Mariano Sigman, and Reinhold Kliegl. Working memory differences in long-distance dependency resolution. *Frontiers in Psychology*, pages 1–16, 2015.

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

[8] Marten van Schijndel, Luan Nguyen, and William Schuler. An analysis of memory-based processing costs using incremental deep syntactic dependency parsing. In *Proc. of CMCL 2013*. Association for Computational Linguistics, 2013.

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

Acknowledgments: This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1343012 and NSF EAGER 1551313.