Cue-based retrieval of parsing rules

Jakub Dotlačil

Cue-based retrieval models have been successful in simulating behavioral measures in the resolution of syntactic dependencies, e.g., subject-verb or relative pronoun-verb dependencies ([1], [2], [3], a.o.). In this work, it will be shown that cue-based retrieval can go beyond modelling the resolution of dependencies. It is straightforwardly compatible with a class of parsers studied in computational linguistics (transition-based parsers, see [4]). Combining cue-based retrieval with transition-based parsing leads to novel psycholinguistic parsers, which (i) can be embedded in the cognitive architectures ACT-R, yet are data-driven, not manually coded (unlike previous ACT-R parsers) (ii) make predictions for psycholinguistic data like on-line behavioral measures without any extra stipulated linking function (unlike previous transition-based parsers in computational linguistics) (iii) are conceptually appealing since they provide a single mechanism for the retrieval of syntactic dependencies and for parsing and a single explanation of either cognitive difficulty (in contrast to previous data-driven psycholinguistic parsers, e.g., [5], [6]). The parser is tested on two data sets: Natural Stories corpus [7] and a self-paced reading study of relative clauses [8].

Cue-based retrieval assumes that memory items are content-addressable and that the memory system uses retrieval cues (e.g., *subject*, *plural* for plural subject retrieval) to find relevant items in memory. The activation of an item increases when it is matched by more retrieval cues and when the retrieval cues are more discriminating. An increase in activation increases a chance of retrieval success and decreases retrieval times ([1], among many others). **Transition-based parsing** is a parsing system that predicts transitions from one parsing state to another by finding the correct parsing step, see (1) for a shift-reduce parser for the sentence *John dances*. The parser has information about its context, represented here as $\mathcal{S}$ and $\mathcal{W}$, and chooses an action which leads to a new context (*shift* shifts the leftmost word in $\mathcal{W}$ to the list of trees and assigns a label to it, *reduce* reduces the rightmost tree structure(s) into a novel tree). Parsing is finished when no upcoming word is present and no reduction can be done among trees. Assuming that finding the right parsing step is a case of memory retrieval and the parsing context ($\mathcal{S}$ and $\mathcal{W}$) serves as the list of retrieval cues, we can conceptualize parsing as just a special case of cue-based retrieval. In parallel with other cases of cue-based retrieval, the model predicts cognitive difficulties (increased latencies, decreased accuracies) if only few retrieval cues find match in memory and/or when retrieval cues are not discriminating because they are shared by many items in memory (cue overload).

Testing cue-based parsing: We construct and collect all parsing steps (assuming a shift-reduce parser) with their context in Penn Treebank, up to section 21 ([9]). We assume that these steps+contexts constitute the memory of the parser. When the parser parses a new sentence, it uses the cues from the current context to find the parsing step with the highest activation in its memory. The model predicts that the activation of the retrieved parsing step should negatively correlate with reading times (RTs). We test this on [7]. Using a mixed-effect model, we see that Activation is indeed a significant negative predictor of RT even after accounting for frequency, position, word length and bigram and trigram frequency, see the left table in (2). The negative Activation effect is moreover driven by the number of matching cues between the currently parsed context and the retrieved parsing step, just as cue-based retrieval predicts, see the right table in (2). To show that the approach allows us to provide a single account of parsing and the resolution of dependencies, we consider self-paced reading data from [8], which has been used to model cue-based retrieval for relative pronoun-verb dependencies. We model reading times by connecting activations (from retrieved lexical items, dependents and parsing steps) to latencies using the standard ACT-R formula (see (3) and [1]). After estimating parameters F and f *once for all types of retrieval*, we get a good fit to the data, Fig. 1. The fit is decreased when the parsing component is switched off, which shows that the good fit is (also) driven by the cue-based model of parsing.

(1) 1.Starting position: $\mathcal{S} = []$, $\mathcal{W} = [\langle \text{John, PN} \rangle, \langle \text{dances, V} \rangle]$

2.shift $\mathcal{S} = [\langle [_{\text{PN}}\ \text{John}] \rangle]$, $\mathcal{W} = [\langle \text{dances, V} \rangle\rangle]$

3.reduce (unary) $\mathcal{S} = [\langle [_{\text{NP}}\ [_{\text{PN}}\ \text{John}]] \rangle]$, $\mathcal{W} = [\langle \text{dances, V} \rangle]$

4.shift $\mathcal{S} = [\langle [_{\text{NP}}\ [_{\text{PN}}\ \text{John}]] \rangle, \langle [_{\text{V}}\ \text{dances}] \rangle]$

5.reduce (unary) $\mathcal{S} = [\langle [_{\text{NP}}\ [_{\text{PN}}\ \text{John}]] \rangle, \langle [_{\text{VP}}\ [_{\text{V}}\ \text{dances}]] \rangle]$

6.reduce (binary) $\mathcal{S} = [\langle [_{\text{S}}\ [_{\text{NP}}\ [_{\text{PN}}\ \text{John}]]\ [_{\text{VP}}\ [_{\text{V}}\ \text{dances}]]] \rangle]$

(2)

	Estimate	t-value
Position	0.034	1.87
Word length	10.75	16.06
Log(freq)	-0.26	-1.95
Length:Log(freq)	-0.53	-13.56
Log(bigram)	-0.004	-0.02
Log(trigram)	-0.56	-2.64
Activation	**-0.14**	**-2.04**

	Estimate	t-value
Position	0.02	1.09
Word length	11.24	18.60
Log(freq)	-0.40	-2.98
Length:Log(freq)	-0.56	-15.96
Log(bigram)	-0.20	-1.21
Log(trigram)	-1.00	-7.04
Number of matching cues	**-0.29**	**-5.04**

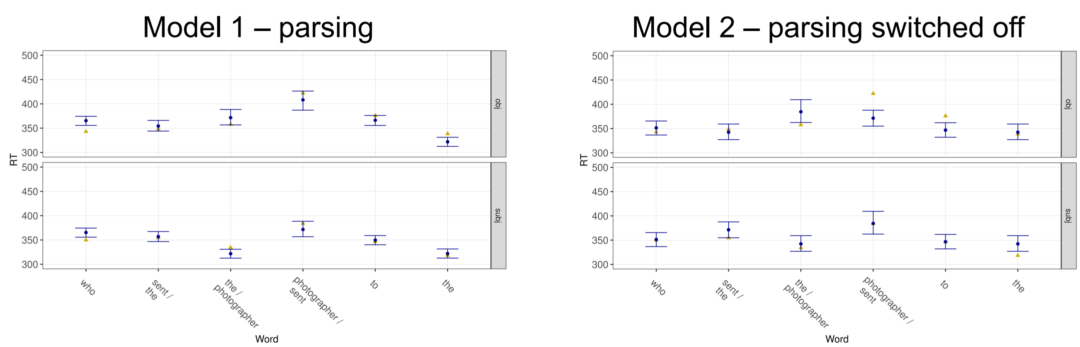

Figure 1: Modelling object-relative and subject-relative self-paced reading data from [8]. The left graphs show the predictions of the model with parsing. The right graphs show the predictions of the model without parsing. The blue dots are predicted mean RTs. The bars provide the 95% credible intervals. The yellow triangles are observed mean RTs.

(3) $T_i = Fe^{-f \cdot A_i}$ (A - activation of item i; F, f – free parameters)

[1] Lewis et al. 2005. An activation-based model of sentence processing as skilled memory retrieval. *CogSci* 29:1–45. **[2]** Dillon et al. 2013. Contrasting intrusion profiles for agreement and anaphora. *JML* 69:85–103. **[3]** Jäger et al. 2017. Similarity-based interference in sentence comprehension. *JML* 94:316–339. **[4]** Nivre. 2004. Incrementality in deterministic dependency parsing. *Workshop on Incremental Parsing*, 50-57. **[5]** Hale. 2001. A probabilistic Earley parser as a psycholinguistic model. The 2nd Meeting of the NAACL, 159–166. **[6]** Boston et al. 2011. Parallel processing and sentence comprehension difficulty. *LCP* 26:301–349. **[7]** Futrell et al. 2018. The natural stories corpus. *LREC 2018*, 76–82. **[8]** Grodner et al. 2005. Consequences of the serial nature of linguistic input for sentenial complexity. *CogSci* 29:261–291. **[9]** Marcus et al. 1993. Building a large annotated corpus of English. *CompLing* 19:313–330.