

Lecture 2, 09/03/2019

Motivation for Tree Adjoining Grammars: introduction to sentential structure

Why should we have formal explicit models of language structure , especially in an age of deep learning and learned representations?

Intuition suggests that such structure exists.

We may want to recover this structure to pass down downstream applications.

Inductive bias (?): Limit the search space.

You should know what your options are, even if you choose *not* to use such models.

What are some language characteristics we should try to capture?

1. The president put \$40 billion into department A's budget but only \$40 into department B's
2. cashiers put baskets in boxes
3. cashiers put boxes in baskets
4. cashiers put boxes and baskets that had lovely bows and were practical -> [baskets that had lovely bows and were practical] is like baskets
5. * cashiers put boxes in put -> only certain types of things can be put in certain positions
6. * cashiers put baskets
7. * cashiers put in boxes

8. cashiers put baskets in boxes ->

[[cashiers]_{noun phrase (NP), or subject}

[[put]_{v, or main verb}

[baskets]_{NP, or direct object}

[[in]_{preposition} [boxes]_{NP}]_{prepositional phrase (PP), or location}

]verb phrase (VP), or predicate

]S, or sentence; "main word" is the predicate's verb

Can we reuse a pre-existing, well-known, efficient formalism?

Example context-free grammar (CFG) fragment

Finite set of categories (e.g., VP), including distinguished start symbol S, finite disjoint set of terminals (e.g. "put"), finite set of decomposition rules, each with the left hand side = exactly one nonterminal.

9. VP → V NP PP (decomposable categories are uppercase by convention)
10. V → put (terminals are in lowercase by convention)
11. V → destroy

Parse trees are *induced*, or the parse trees themselves induce the sentence

Handling local restrictions (let's be clever engineers)

Lexical information (characteristics of individual words, or *lexical items*) is important.

12. **she** puts boxes in baskets versus * **cashiers** puts boxes in baskets
13. **they** put boxes in baskets versus * **cashiers** put **they** in baskets -> case mismatch ("subject" v. "direct object")
14. ?? **cashiers** put **sleep** in baskets
15. Lexical entry for "put" includes: subcategorization is (1: NP that is "puttable" 2: PP that is a location)
16. VP → V NP PP
 - VP agreement = V agreement
 - V's subcat 1 = NP
 - V's subcat 2 = PP

(using "=" loosely as "is consistent with"/"unifies with")

Handling gaps in question inversion

17. **what** do the cashiers put in the baskets?
18. **where** do the cashiers put the boxes?
19. * **the cashiers** put in the baskets -> if there's a gap, there must be a filler
20. * **what** do the cashiers put the boxes -> the WH "filler" has to match the gap's characteristics

21. $S_{\text{question}} \rightarrow NP_{\text{wh}} \text{AUX NP VP}$
 - AUX agreement = NP agreement
 - VP gapinfo = NP_{wh}gapinfo
 - VP has an NP gap
 - NP_{wh} is not null
22. NP -> <null trace>
 - NP must have nondefault gapinfo

Add to 16:

- VP gapinfo = NP gapinfo
- PP no gap allowed