Dependency Parsing

CS 485, Fall 2023
Applications of Natural Language Processing
https://people.cs.umass.edu/~brenocon/cs485_f23/

Brendan O'Connor
College of Information and Computer Sciences
University of Massachusetts Amherst
CFG issue

- Substitutability is too strong (e.g. “she” as subject vs object)

Figure 11.5: A grammar that allows *she* to take the object position wastes probability mass on ungrammatical sentences.

Figure 11.6: The left parse is preferable because of the conjunction of phrases headed by France and Italy. More fine-grained NP and VP categories might allow us to make attachment decisions more accurately.

Semantic preferences

In addition to grammatical constraints such as case marking, we have semantic preferences: for example, that conjoined entities should be similar. In Figure 11.6, you probably prefer the left parse, which conjoins France and Italy, rather than the right parse, which conjoins wine and Italy. But it is impossible for a PCFG to distinguish these parses! They contain exactly the same productions, so the resulting probabilities will be the same, no matter how you define the probabilities of each production.
CFG issue

- Substitutability is too strong (PP attachment ambiguity)

Figure 10.1: Two derivations of the same sentence, shown as both parse trees and bracketings

Head rules

• Idea: Every phrase has a head word, that is the "core" or "nucleus" determining its syntactic role

• Head rules: for every nonterminal in tree, choose one of its children to be its “head”. This will define head words.

• Every nonterminal type has a different head rule; e.g. from Collins (1997):

  • If parent is NP,
    • Search from right-to-left for first child that’s NN, NNP, NNPS, NNS, NX, JJR
    • Else: search left-to-right for first child which is NP
Heads in constits.

```
S
  NP
    DT the
    NN lawyer
  VP
    Vt questioned
    NP
      DT the
      NN witness
```
Heads in constits.

```
S
/   |
|    |
NP   VP
|    |
DT   NN  Vt   NP
|    |
the  lawyer  questioned  the  witness

↓

S(questionsed)
/   |
|    |
NP   VP
|    |
NN(questionsed)  NN(witness)
|    |
lawyer  the  witness
```

Adding Headwords to Trees

NP

S

VP

NP

Vt

NP

DT

the

NN

lawyer

questioned

NP

Vt

NP

DT

the

NN

witness

Heads in constits.
Lexicalized CFGs

<table>
<thead>
<tr>
<th>Non-terminal</th>
<th>Direction</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>right</td>
<td>VP SBAR ADJP UCP NP</td>
</tr>
<tr>
<td>VP</td>
<td>left</td>
<td>VBD VBN MD VBZ TO VB VP VBG VBP ADJP NP</td>
</tr>
<tr>
<td>NP</td>
<td>right</td>
<td>N* EX $ CD QP PRP . . .</td>
</tr>
<tr>
<td>PP</td>
<td>left</td>
<td>IN TO FW</td>
</tr>
</tbody>
</table>

Table 11.3: A fragment of head percolation rules

Figure 11.9: Lexicalization can address ambiguity on coordination scope (upper) and PP attachment (lower)
From constituency structure to dependency graphs

Figure 11.1: Dependency grammar is closely linked to lexicalized context free grammars: each lexical head has a dependency path to every other word in the constituent. (This example is based on the lexicalization rules from § 10.5.2, which make the preposition the head of a prepositional phrase. In the more contemporary Universal Dependencies annotations, the head of with claws would be claws, so there would be an edge scratch → claws.)
• Dependencies tend to be less specific than constituent structure

(a) Flat

(b) Two-level (PTB-style)

(c) Chomsky adjunction

(d) Dependency representation

Figure 12.3: The three different CFG analyses of this verb phrase all correspond to a single dependency structure. So if you didn't think there was any meaningful difference between these three constituent representations, you may view this as an advantage of the dependency representation.

Dependency grammar still leaves open some tricky representational decisions. For example, coordination is a challenge: in the sentence, Abigail and Max like kimchi (Figure 12.4), which word is the immediate dependent of the main verb likes? Choosing either Abigail or Max seems arbitrary; for fairness we might choose and, but this seems in some ways to be the least important word in the noun phrase. One typical solution is to simply choose the left-most item in the coordinated structure — in this case, Abigail.

Another alternative, as shown in Figure 12.4c, is a collapsed dependency grammar in which conjunctions are not included as nodes in the graph, but are instead used to label the edges (De Marneffe et al., 2006). Popel et al. (2013) survey alternatives for handling this phenomenon across several dependency treebanks.

The same logic that makes us reluctant to accept and as the head of a coordinated noun phrase may also make us reluctant to accept a preposition as the head of a prepositional phrase. In the sentence cats scratch people with claws, surely the word claws is more central than the word with — and it is precisely the bilexical relations between scratch, claws, and people that help guide us to the correct syntactic interpretation. Yet there are also arguments for preferring the preposition as the head — as we saw in section 11.5, the preposition itself is what helps us to choose verb attachment in meet the President on Monday and noun attachment in meet the President of Mexico.
Headedness for *phrase* relations

- Is a given word $X$ the subject of verb $Y$?
- Is a given *phrase* $X$ the subject of verb $Y$?
Universal Dependencies

• Dependency treebanks are available for many different languages
  • [https://universaldependencies.org/](https://universaldependencies.org/)

• Many open-source dependency parsers (and tagging/POS/morphology) trained on them are also widely available; e.g. Stanza, SpaCy, etc.
  • They typically directly predict dependencies with another parsing algorithm (shift-reduce, not CKY)
Dependency applications

- Dependencies can be used as less sparse alternative to n-grams
- Sometimes helps, sometimes doesn’t
- Dependency relations can be selected for semantic relationships...
Figure 12.8: Google n-grams results for the bigram *write code* and the dependency arc *write => code* (and their morphological variants)

- **Goldberg & Orwant 2013**: historical dependencies from google books ([https://books.google.com/ngrams/](https://books.google.com/ngrams/))
Dependency pattern statistics

4.3.1 IS_A

The IS_A relation covers any nominal or adjectival properties stated to directly pertain to the target entity, represented using the following patterns:

1. target \( \overset{nsubj}{} \rightarrow \text{property}_{nom} \)
2. \( \text{property}_{adj} \overset{nsubj}{} \rightarrow \text{target} \)
3. \( \text{target} \overset{apos}{} \rightarrow \text{property}_{nom} \)
4. \( \text{target} \overset{compound}{} \rightarrow \text{property}_{nom} \)
5. \( \text{target} \overset{amod}{} \rightarrow \text{property}_{adj} \)
6. \( \text{target} \overset{nsubj}{} \rightarrow \text{property}_{nom} \overset{amod}{} \rightarrow \text{property}_{adj} \)
7. \( \text{target} \overset{apos}{} \rightarrow \text{property}_{nom} \overset{amod}{} \rightarrow \text{property}_{adj} \)
<table>
<thead>
<tr>
<th>Relation</th>
<th>Trump-Leaning ($t &lt; -2$)</th>
<th>Biden-Leaning ($t &gt; 2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_A(fauci, $\text{property}_{\text{nom}}$)</td>
<td>murderer**, joke**, hack**, fraud*, rat*, flip*, idiot, flop, state, prison, fake, jail</td>
<td>nih**, hero, md, director, president</td>
</tr>
<tr>
<td>IS_A(fauci, $\text{property}_{\text{adj}}$)</td>
<td>fake*, little*, deep, liberal, wrong, corrupt</td>
<td>beloved, optimistic, best</td>
</tr>
<tr>
<td>AS_AGENT(fauci, verb)</td>
<td>sweat**, force**, need*, help*, read*, lie*, know*, let*, not_fund*, not_understand*, flip, predict, write, make, stick, hold, prove, want, not_say, admit, not_get, demand, issue, laugh, state, put, spread, pull</td>
<td>speak**, join*, warn*, throw, not_recommend, offer, provide, respond, consider, debunk, fail, reveal</td>
</tr>
<tr>
<td>AS_PATIENT(fauci, verb)</td>
<td>not_trust***, screw, prosecute, grill, keep to, arrest, expose, lock, do to, remove, accord to, look like, mean, blast, read</td>
<td>know*, feature, discredit, threaten, worship, join, insult</td>
</tr>
<tr>
<td>HAS_A(fauci, object)</td>
<td>friend*, nih*, family, mind, hand, ex-employee, involvement, fraud, mask</td>
<td>guidance, time</td>
</tr>
<tr>
<td>AS_CONJUNCT(fauci, conj.)</td>
<td>gates***, obama**, bill gates*, biden*, brix, cdc, rest, covid, nih, company, government</td>
<td>director, experts</td>
</tr>
</tbody>
</table>

Table 5: TweetIE extractions with at least 20 unique users with a county-level political valence $t$-statistic outside of [-2, 2]. Results are reported in decreasing absolute value $t$-statistic. * $|t| > 3$, ** $|t| > 4$, *** $|t| > 5$.

- From geo-located tweets, Mar-Dec 2020

[O'Conner, 2022]
Dependency paths

They had previously bought bighorn sheep from Comstock.

The paths extracted from this sentence and their meanings are:

(a) N:subj:V←buy→V:from:N
   ≡ X buys something from Y
(b) N:subj:V←buy→V:obj:N
   ≡ X buys Y
(c) N:subj:V←buy→V:obj:N→sheep→N:nn:N
   ≡ X buys Y sheep
   ≡ X sheep is bought from Y
(e) N:obj:V←buy→V:from:N
   ≡ X is bought from Y

An inverse path is also added for each one above.

● Dependency path corresponds to a lexico-syntactic pattern

[Lin and Pantel 2001]
Distributional similarity

- “You shall know a word by the company it keeps” [Firth, 1957]

- Simple single-word (lexical semantics) example: “duty” vs “responsibility”

adj. modification, verbs they’re arguments of?

- duty can be modified by adjectives such as additional, administrative, assigned, assumed, collective, congressional, constitutional, …, so can responsibility;

- duty can be the object of verbs such as accept, articulate, assert, assign, assume, attend to, avoid, become, breach, …, so can responsibility.