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.

Subsumption
There are several choices for annotating PP attachment (c) Jacob Eisenstein 2014-2017. Work in progress.
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 (c) Jacob Eisenstein 2014-2017. Work in progress.
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.
Heads in constits.

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

↓

S(questioned)
  NP(lawyer)  VP(questioned)
  |         |     |
  DT the  NN lawyer Vt questioned  NP witness
  the lawyer questioned the witness
Lexicalized CFGs

Table 11.3: A fragment of head percolation rules

<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>

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
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/

• 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...
12.4. APPLICATIONS

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/)
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. \( \text{target} \xleftarrow{\text{nsubj}} \text{property}_{\text{nom}} \)
2. \( \text{property}_{\text{adj}} \xrightarrow{\text{nsubj}} \text{target} \)
3. \( \text{target} \xrightarrow{\text{appos}} \text{property}_{\text{nom}} \)
4. \( \text{target} \xrightarrow{\text{compound}} \text{property}_{\text{nom}} \)
5. \( \text{target} \xrightarrow{\text{amod}} \text{property}_{\text{adj}} \)
6. \( \text{target} \xleftarrow{\text{nsubj}} \text{property}_{\text{nom}} \xrightarrow{\text{amod}} \text{property}_{\text{adj}} \)
7. \( \text{target} \xleftarrow{\text{appos}} \text{property}_{\text{nom}} \xrightarrow{\text{amod}} \text{property}_{\text{adj}} \)
We obtain 75,325 tweets, which have an electoral words' frequencies greatly vary, rare terms tend we lowercase and normalize the normalization by standard error helps control for false discovery rate or other methods could be applied to be sentiment average outliers; the continuous variable of political sentiment. Since a rough filter for traditional statistical significance. than the corpus population's. We require ical sentiment of the tuple is significantly different to determine if the mean author-geography polit-
tic to 2010 margin average of 22.8 and standard deviation of 5.2 Results and Qualitative Evaluation

<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, property&lt;sub&gt;nom&lt;/sub&gt;)</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, property&lt;sub&gt;adj&lt;/sub&gt;)</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, de-bunk, 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

[ Eggleston and O'Connor, 2022]
Dependency paths

They had previously bought bighorn sheep from Comstock.

The paths extracted from this sentence and their meanings are:

(a) $N: subj \leftarrow V \rightarrow V: from: N$
   
   $\equiv X$ buys something from $Y$

(b) $N: subj \leftarrow V \rightarrow V: obj: N$
   
   $\equiv X$ buys $Y$

(c) $N: subj \leftarrow V \rightarrow V: obj: N \rightarrow sheep \rightarrow N: nn: N$
   
   $\equiv X$ buys $Y$ sheep

(d) $N: nn: N \leftarrow sheep \leftarrow N: obj \leftarrow V \rightarrow V: from: N$
   
   $\equiv X$ sheep is bought from $Y$

(e) $N: obj \leftarrow V \rightarrow V: from: N$
   
   $\equiv X$ is bought from $Y$

An inverse path is also added for each one above.

- Dep path corresponds to a lexico-syntactic pattern
- Dep path is a chain of relation conjunctions, leaving further modifications unspecified
- Which dep paths to get? Heuristics to alleviate sparsity (L&P require content words, limit path length, etc.)
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*. 