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)

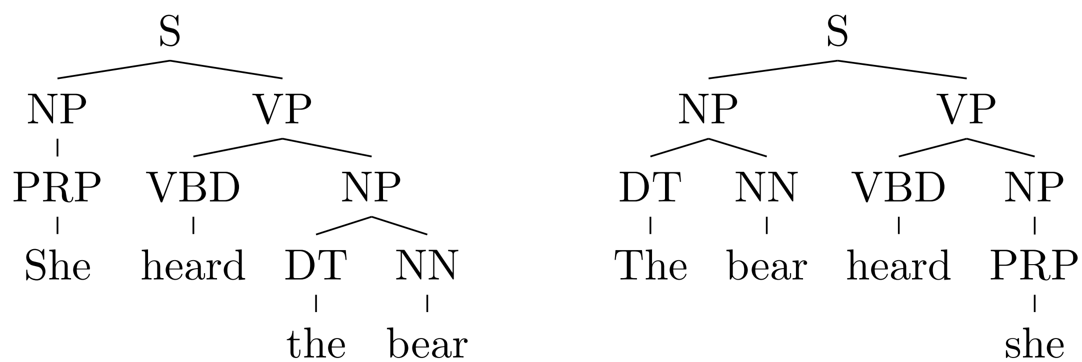
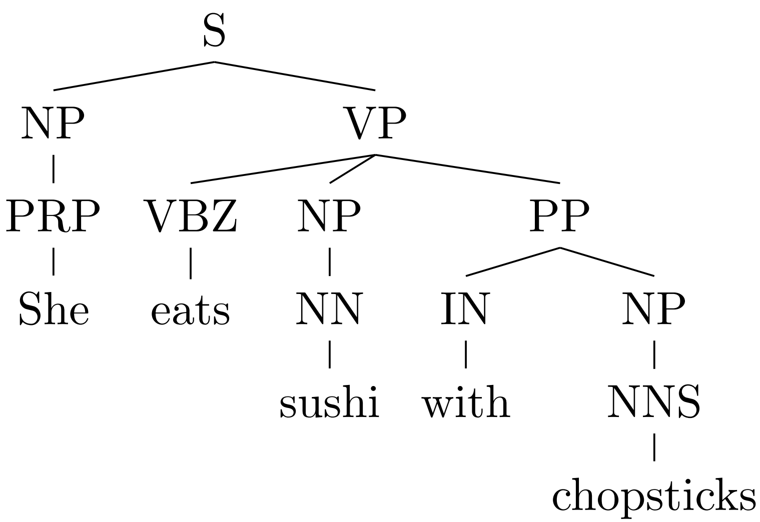


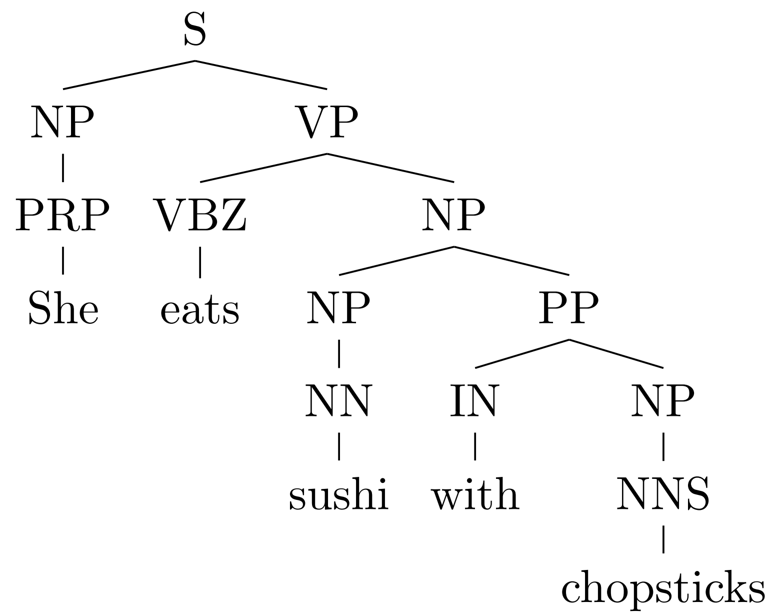
Figure 11.5: A grammar that allows *she* to take the object position wastes probability mass on ungrammatical sentences.

[From Eisenstein (2017)]

CFG issue

- Substitutability is too strong (PP attachment ambiguity)



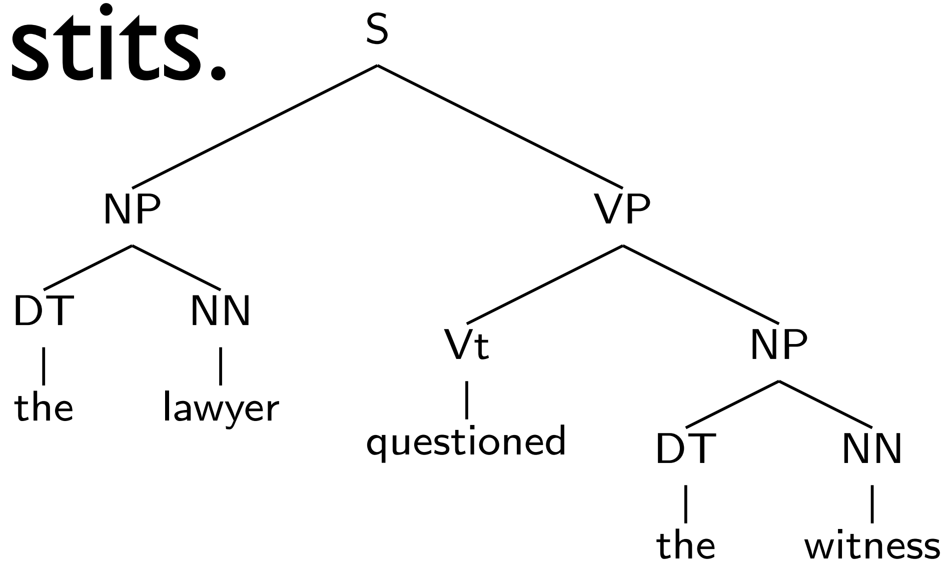


Head rules

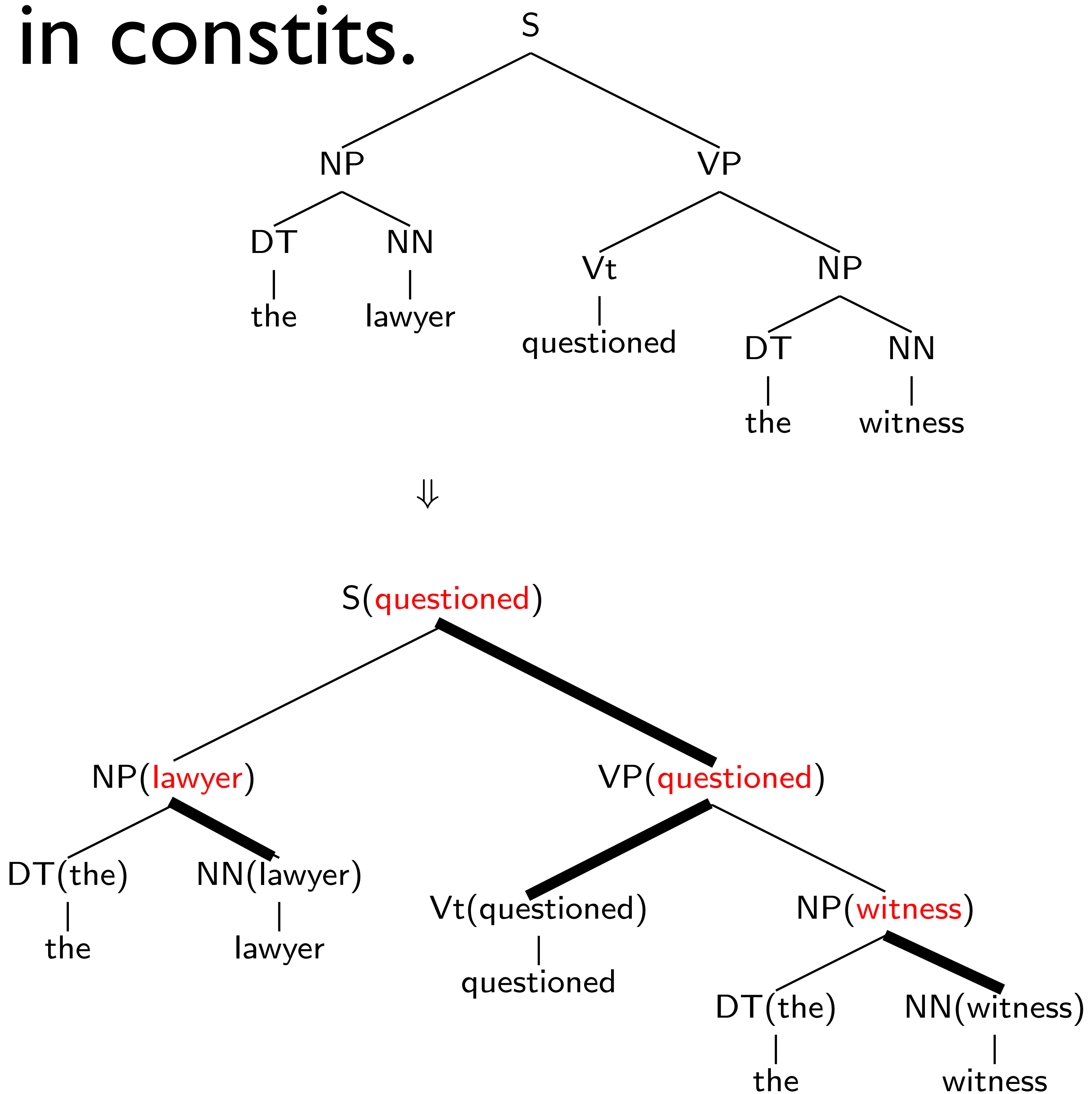
- Idea: Every phrase has a *head word*, that is the "core" or "nucleus" determining its syntactic role
- Head rules: for every nonterminal in tree, choose one of its children to be its “head”. This will define head words.
- Every nonterminal type has a different head rule; e.g. from Collins (1997):

> - 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.



Lexicalized CFGs

Non-terminal	Direction	Priority
S	right	VP SBAR ADJP UCP NP
VP	left	VBD VBN MD VBZ TO VB VP VBG VBP ADJP NP
NP	right	N* EX $ CD QP PRP . . .
PP	left	IN TO FW

Table 11.3: A fragment of head percolation rules

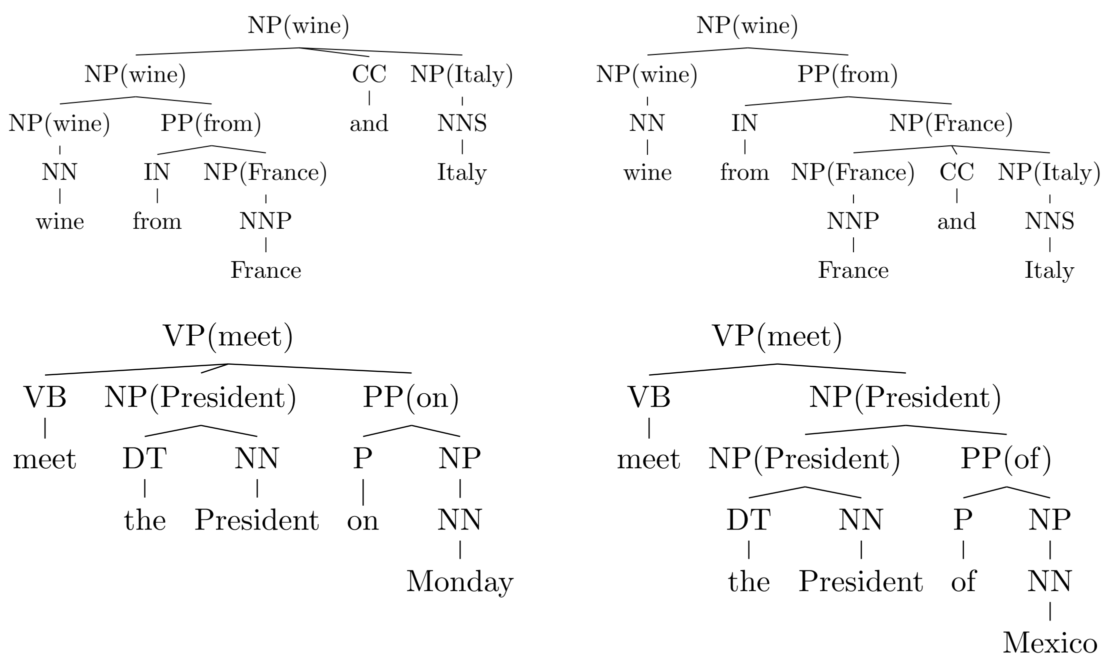
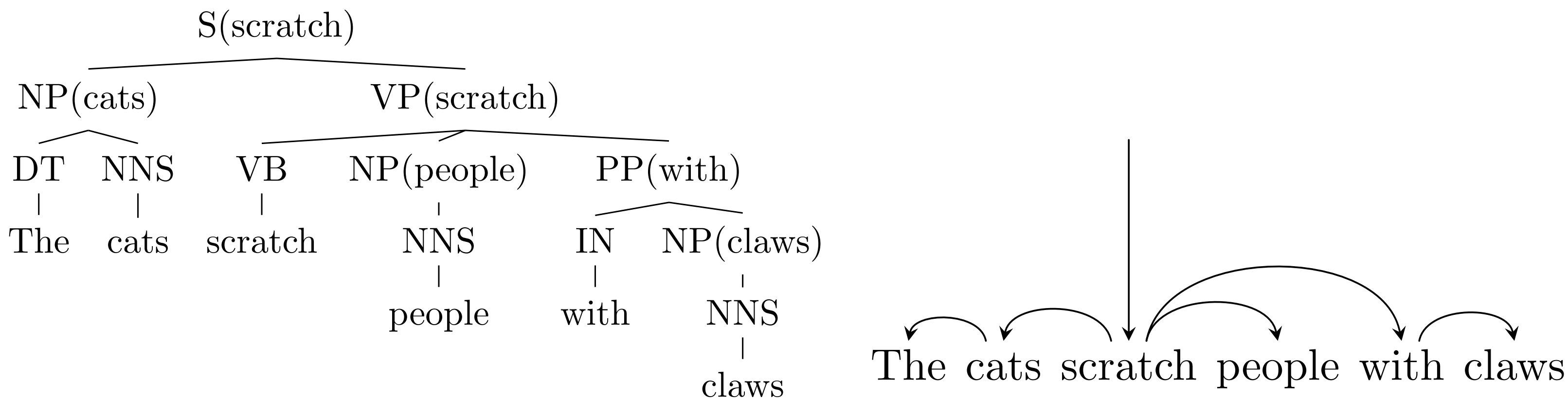


Figure 11.9: Lexicalization can address ambiguity on coordination scope (upper) and PP attachment (lower)

[From Eisenstein (2017)]

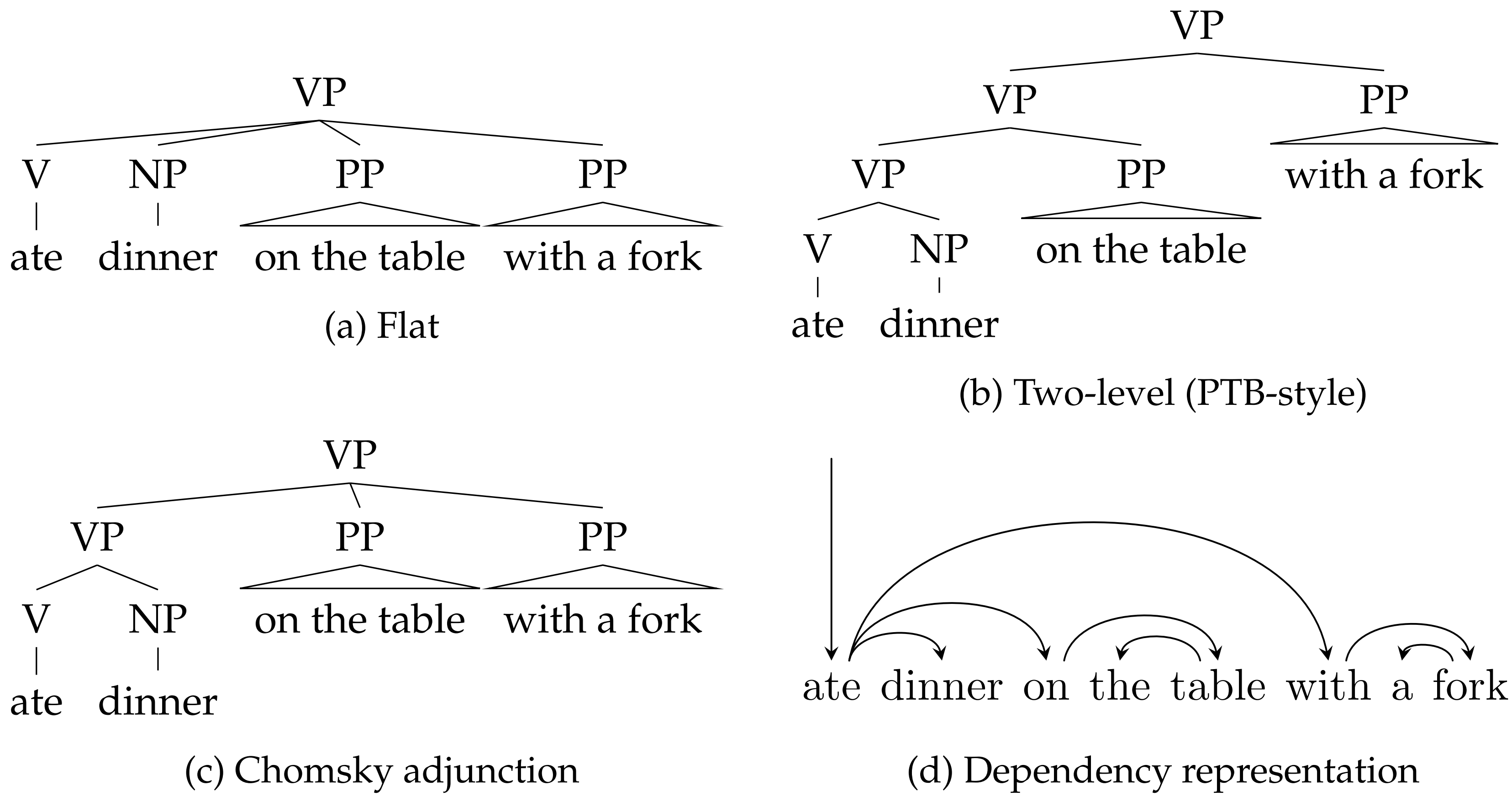
From constituency structure to dependency graphs



(a) lexicalized constituency parse (b) unlabeled dependency tree

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

[Example: Jacob Eisenstein]

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...

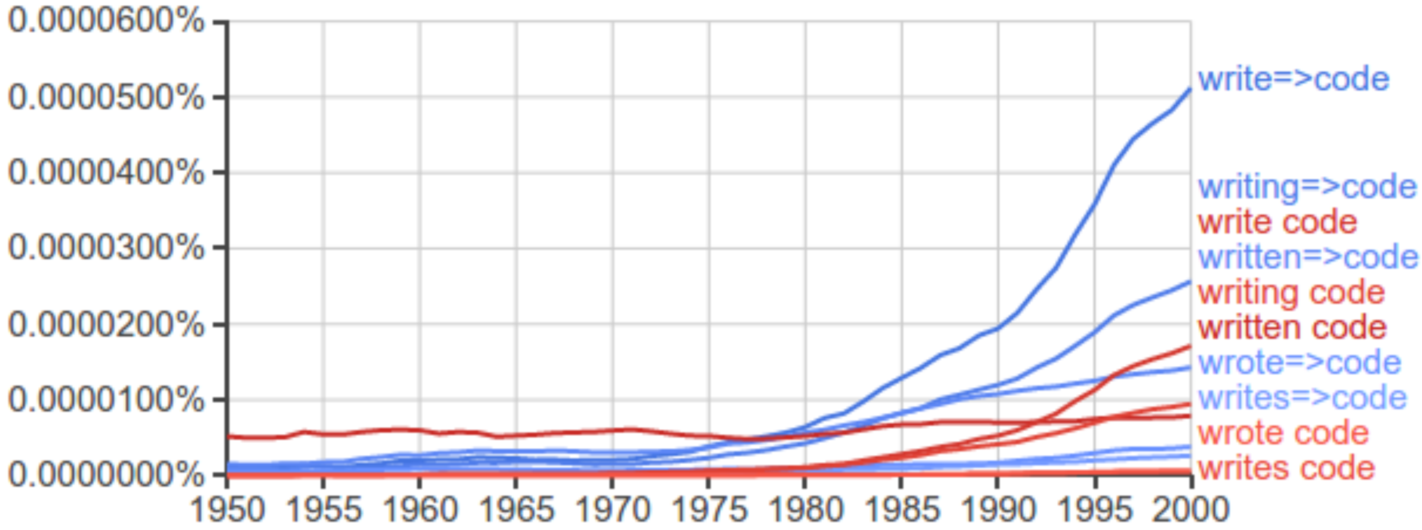


Figure 12.8: Google n-grams results for the bigram *write code* and the dependency arc *write => code* (and their morphological variants)

- *Goldberg & Orwant 2013*: historical dependencies from google books (https://books.google.com/ngrams/)

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:[5]

1. target $\xleftrightarrow{\text{nsubj}}$ property$_{nom}$
2. property$_{adj}$ $\xrightarrow{\text{nsubj}}$ target
3. target $\xleftrightarrow{\text{appos}}$ property$_{nom}$
4. target $\xrightarrow{\text{compound}}$ property$_{nom}$
5. target $\xrightarrow{\text{amod}}$ property$_{adj}$
6. target $\xleftrightarrow{\text{nsubj}}$ property$_{nom}$ $\xrightarrow{\text{amod}}$ property$_{adj}$
7. target $\xleftrightarrow{\text{appos}}$ property$_{nom}$ $\xrightarrow{\text{amod}}$ property$_{adj}$

[Eggleston and O'Connor, 2022]

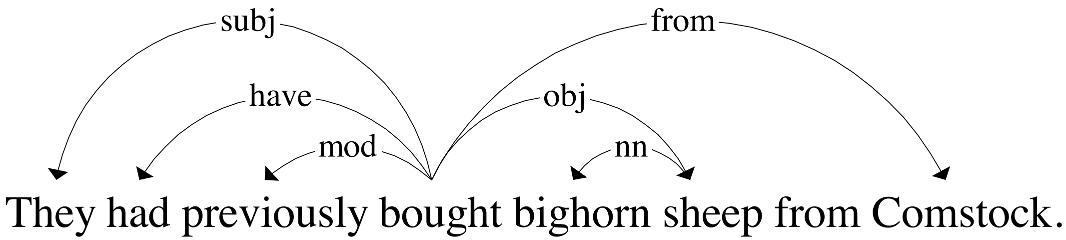
Relation	**Trump-Leaning** ($t < -2$)	**Biden-Leaning** ($t > 2$)
IS_A(fauci, $\textit{property}_{\text{nom}}$)	murderer**, joke**, hack*, fraud*, rat*, flip*, idiot, flop, state, prison, fake, jail	nih**, hero, md, director, president
IS_A(fauci, $\textit{property}_{\text{adj}}$)	fake*, little*, deep, liberal, wrong, corrupt	beloved, optimistic, best
AS_AGENT(fauci, *verb*)	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	speak**, join*, warn*, throw, not_recommend, offer, provide, respond, consider, debunk, fail, reveal
AS_PATIENT(fauci, *verb*)	not_trust***, screw, prosecute, grill, keep to, arrest, expose, lock, do to, remove, accord to, look like, mean, blast, read	know*, feature, discredit, threaten, worship, join, insult
HAS_A(fauci, *object*)	friend*, nih*, family, mind, hand, ex-employee, involvement, fraud, mask	guidance, time
AS_CONJUNCT(fauci, *conj.*)	gates***, obama**, bill gates*, biden*, brix, cdc, rest, covid, nih, company, government	director, experts

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



The paths extracted from this sentence and their meanings are:

(a) N:subj:V←buy→V:from:N
$\equiv$ *X* buys something from *Y*

(b) N:subj:V←buy→V:obj:N
$\equiv$ *X* buys *Y*

(c) N:subj:V←buy→V:obj:N→sheep→N:nn:N
$\equiv$ *X* buys *Y* sheep

(d) N:nn:N←sheep←N:obj:V←buy→V:from:N
$\equiv$ *X* sheep is bought from *Y*

(e) N:obj:V←buy→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) exmaple:
  “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*.