CS11-737 Multilingual NLP

Dependency Parsing

Graham Neubig

Carnegie Mellon University

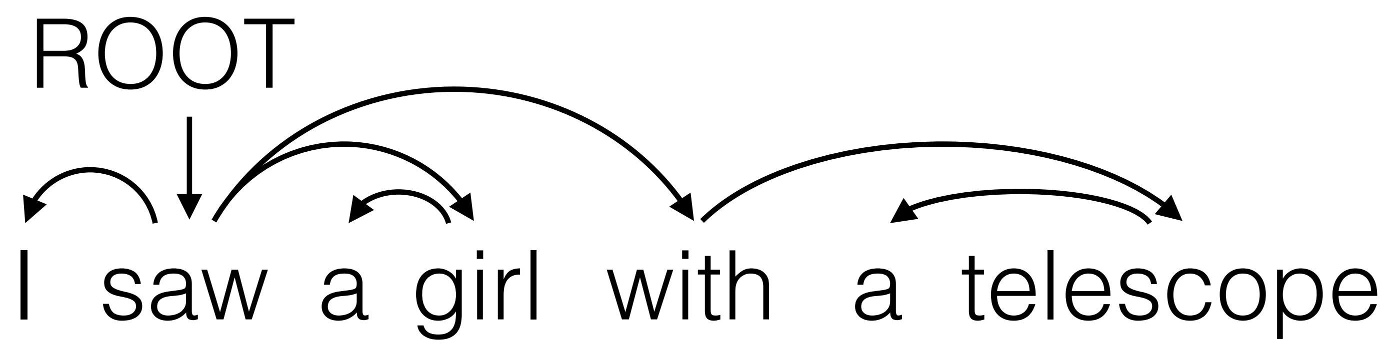
Language Technologies Institute

Site

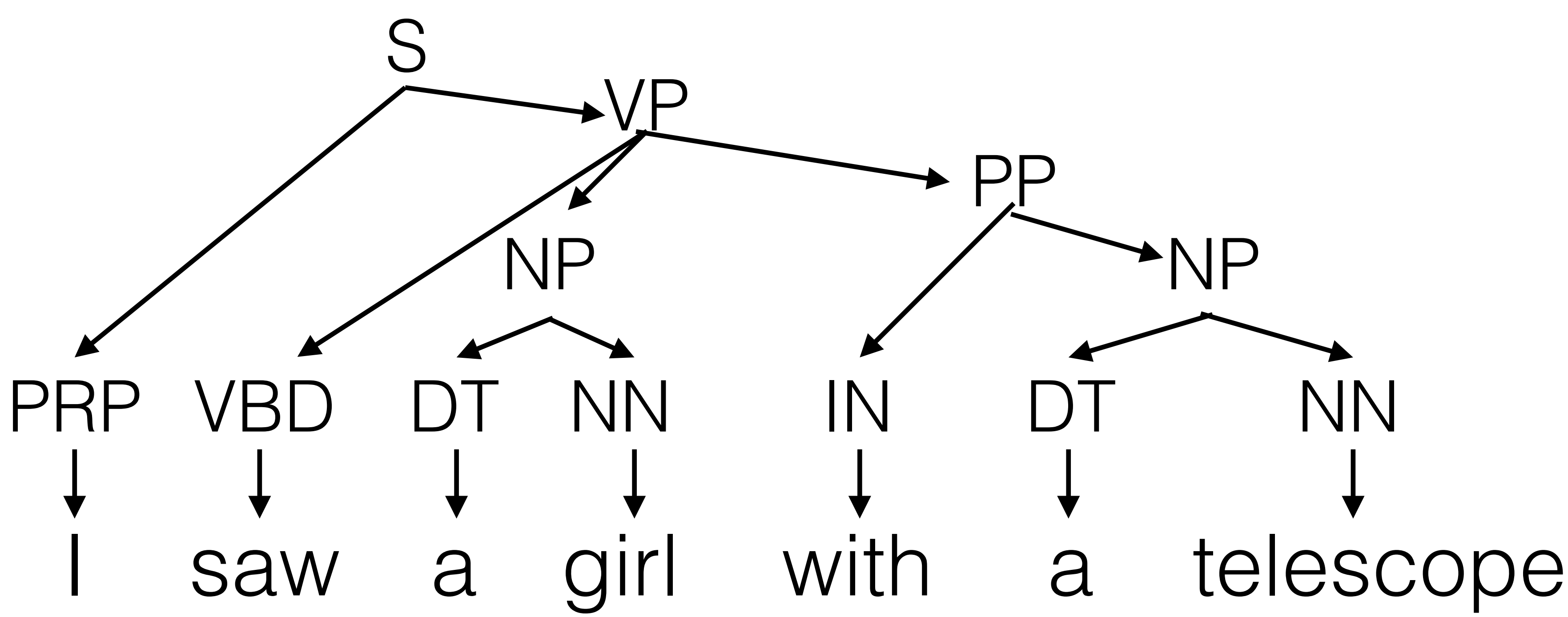
http://phontron.com/class/multiling2022/

Two Types of Linguistic Structure

- **Dependency:** focus on relations between words

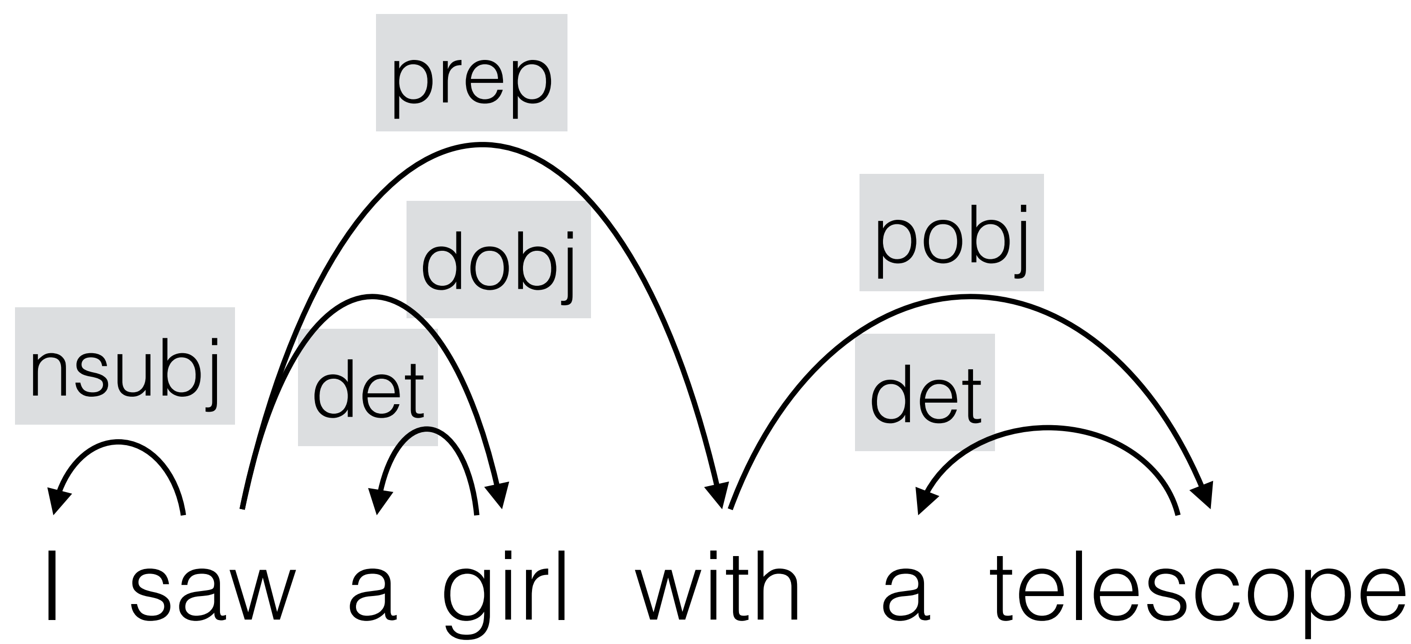


- **Phrase structure:** focus on the structure of the sentence



Why Dependencies?

- Demonstrate the relationships between words in a straightforward way



- Particularly good for multilinguality, e.g. phrase-structure can be hard to define in languages with free word order

Universal Dependencies Treebank

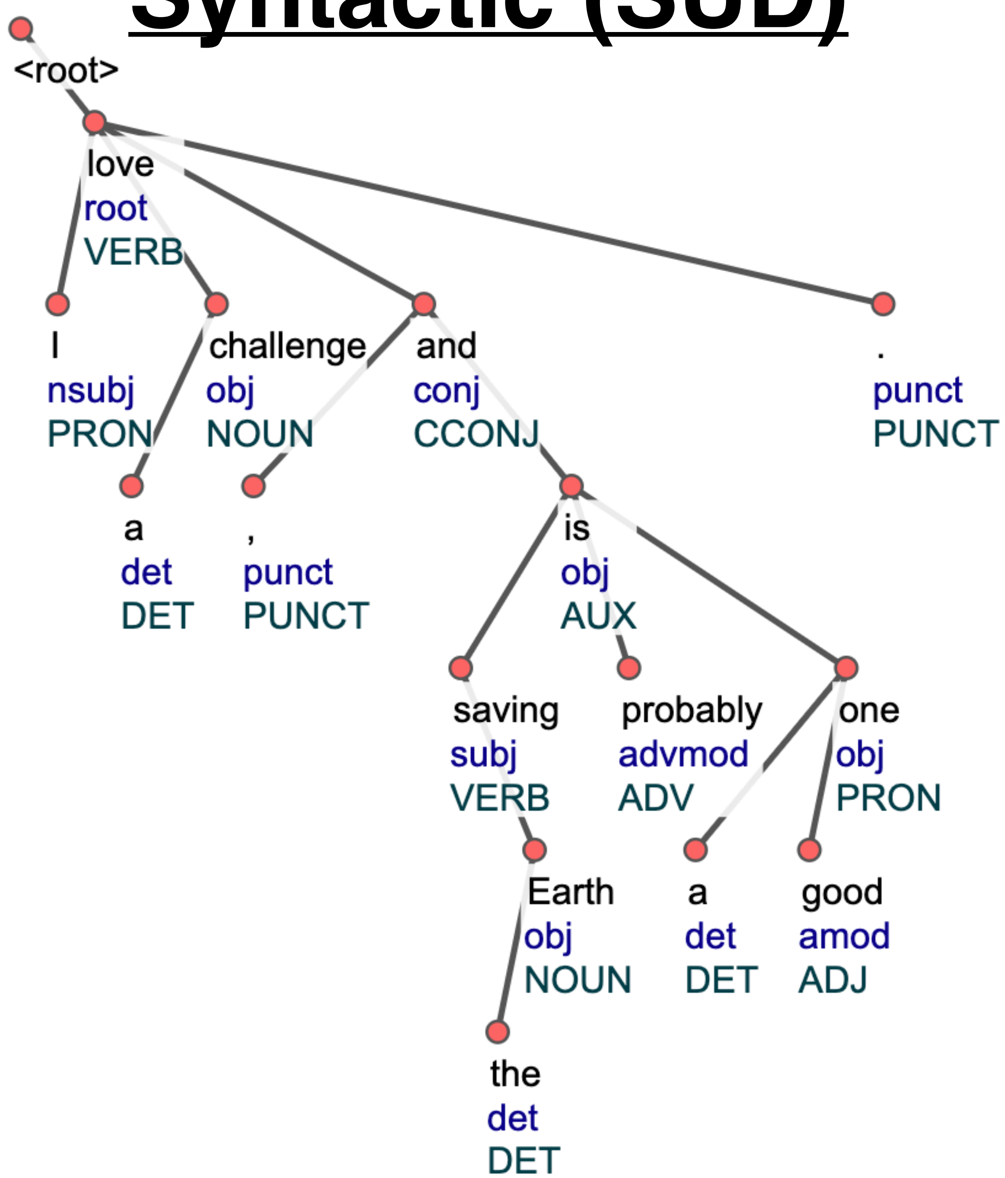
- Standard format for parse trees in many languages

Abaza	1	3K		Northwest Caucasian
Afrikaans	1	49K		IE, Germanic
Akkadian	1	1K		Afro-Asiatic, Semitic
Albanian	1	<1K	W	IE, Albanian
Amharic	1	10K		Afro-Asiatic, Semitic
Ancient Greek	2	416K		IE, Greek
Arabic	3	1,042K	W	Afro-Asiatic, Semitic
Armenian	1	52K		IE, Armenian
Assyrian	1	<1K		Afro-Asiatic, Semitic
Bambara	1	13K		Mande
Basque	1	121K		Basque
Belarusian	1	13K		IE, Slavic
Bhojpuri	2	6K		IE, Indic
Breton	1	10K	W	IE, Celtic
Bulgarian	1	156K		IE, Slavic
Buryat	1	10K		Mongolic
Cantonese	1	13K		Sino-Tibetan

https://universaldependencies.org/

Semantic and Syntactic Dependencies

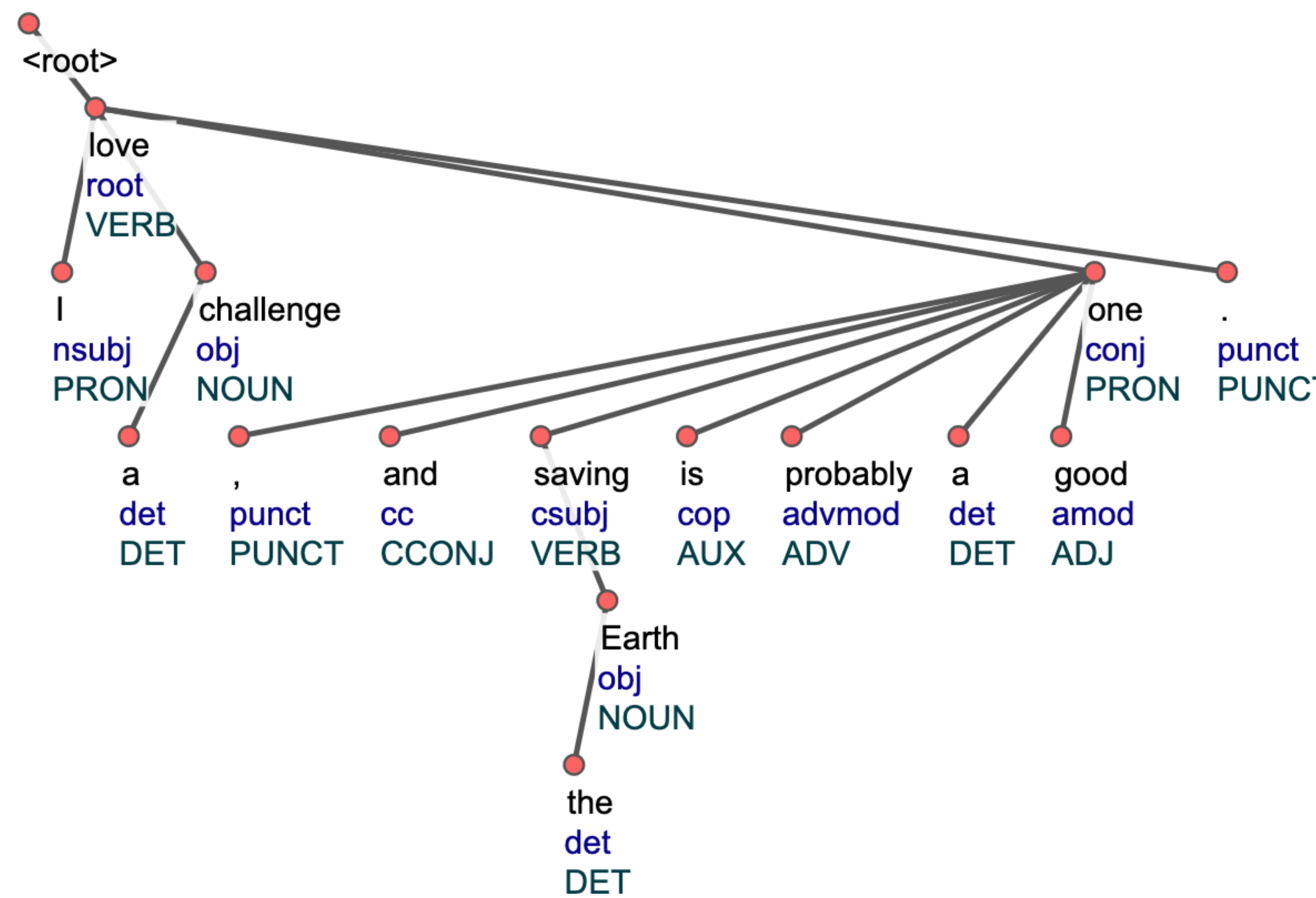
Syntactic (SUD)



Deeper, reflect phrase structure, more function word heads

https://surfacesyntacticud.github.io/

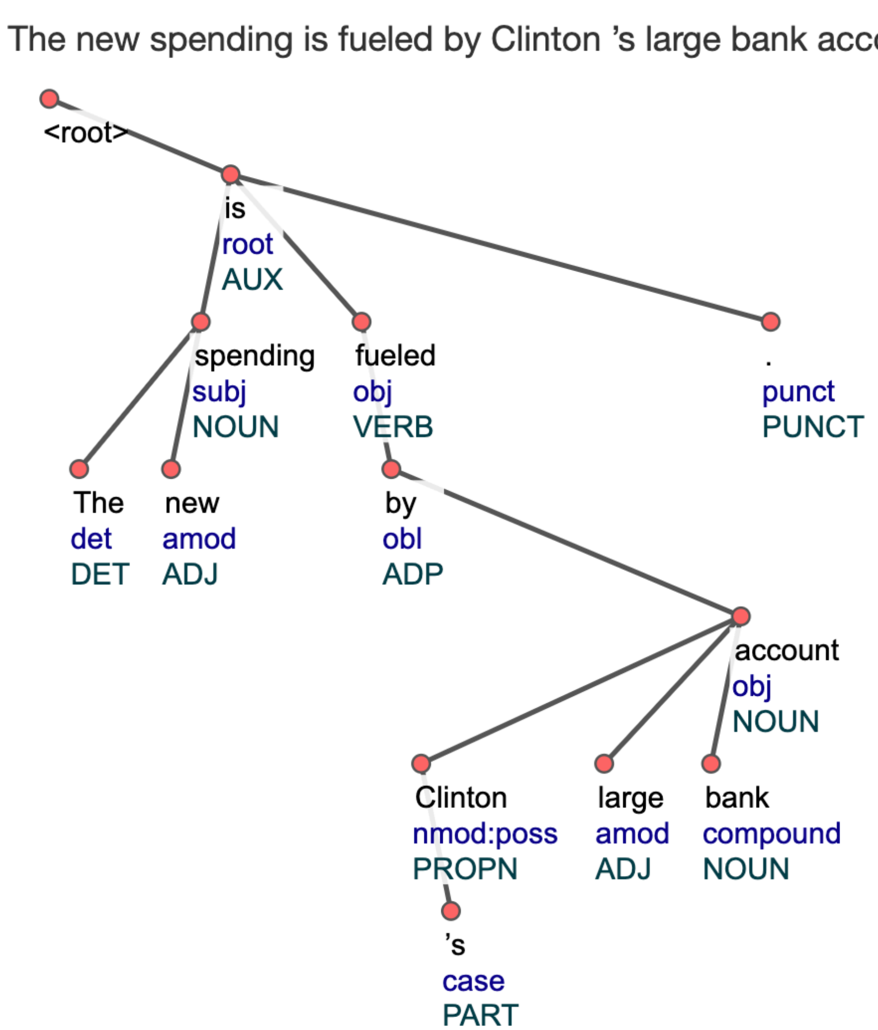
Semantic (UD)



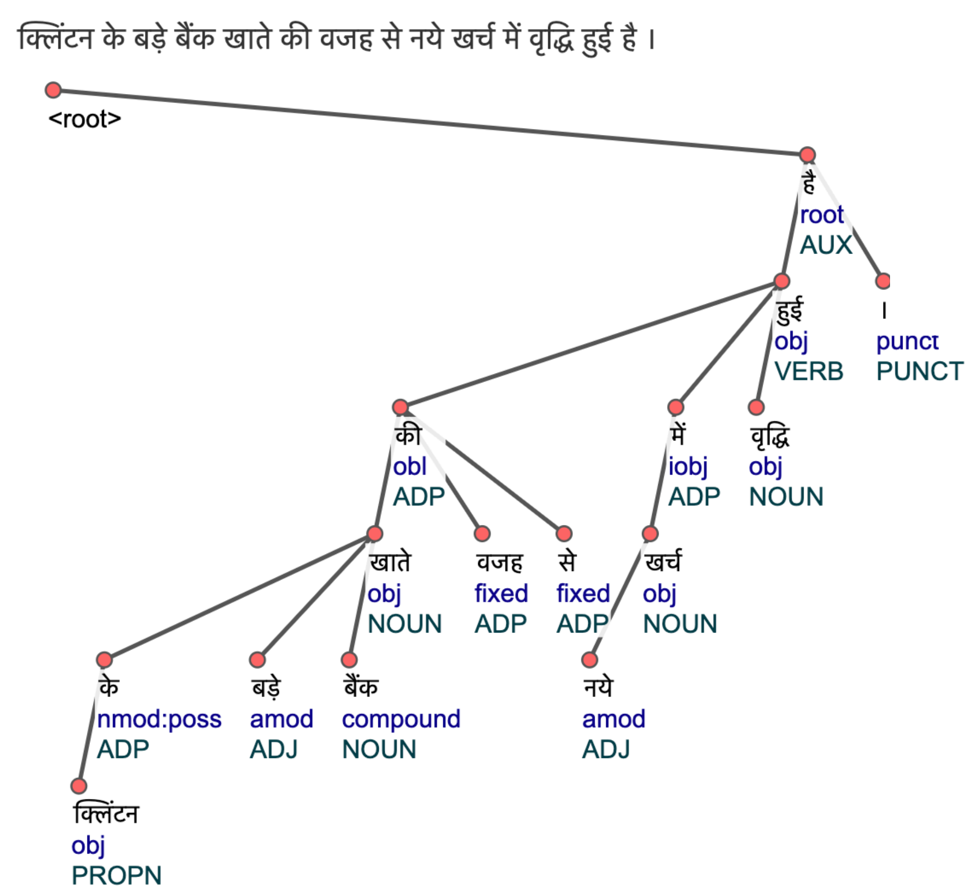
Flatter, semantically related words closer, more content word heads

Cross-lingual Differences In Structure

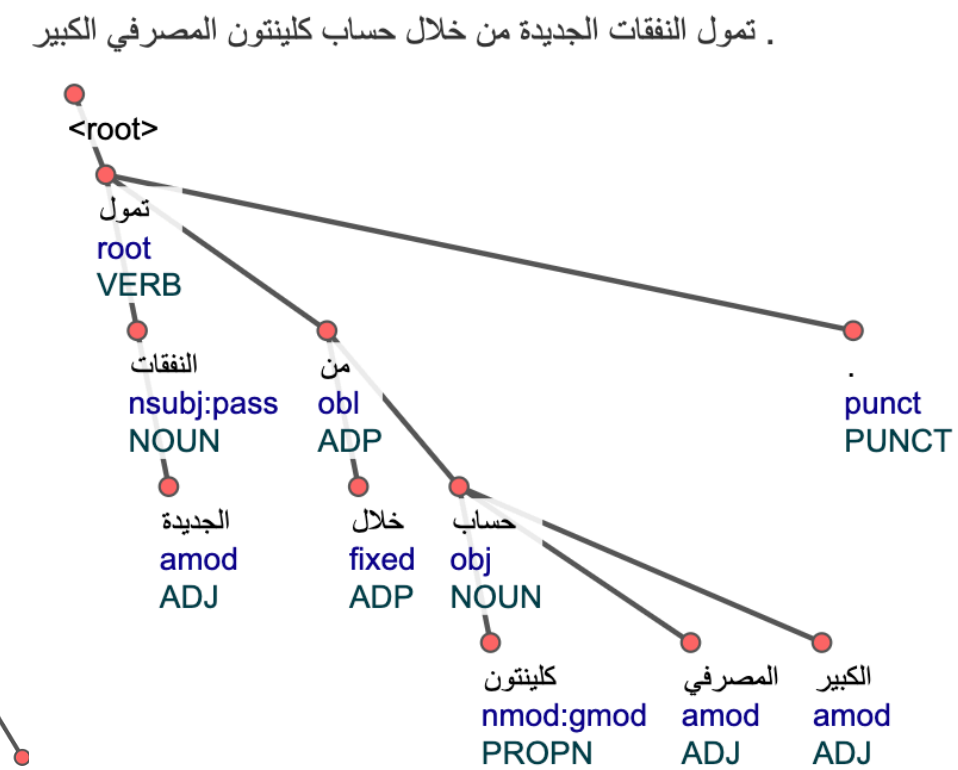
English: SVO



Hindi: Verb Final



Arabic: Verb Initial

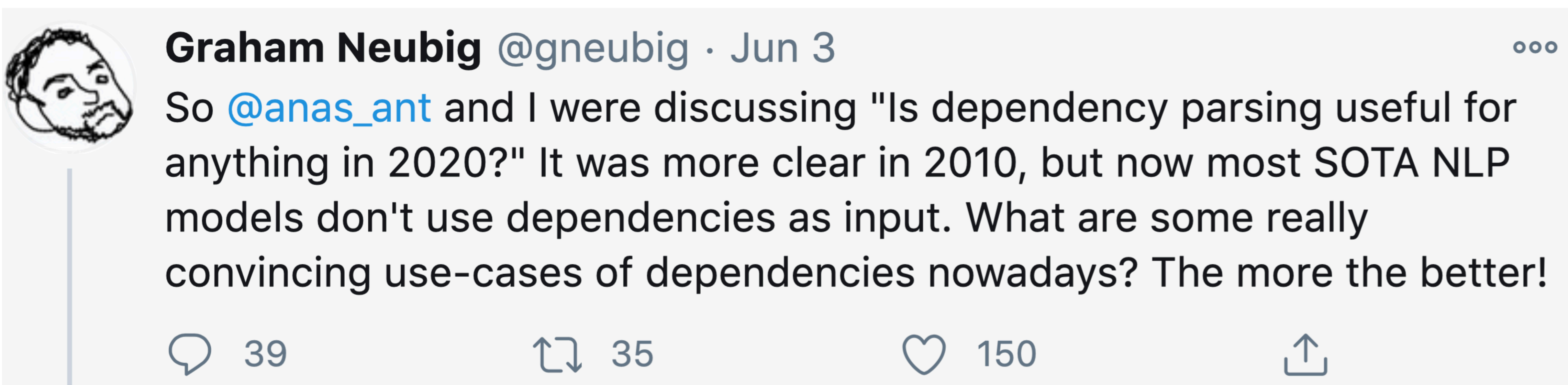


https://universaldependencies.org/treebanks/{en,hi,ar}_pud/index.html

What Can we Do w/ Dependencies?

Use Cases of Dependencies?

- Previously, used for feature engineering in systems (and still useful in some cases)
- Now: more useful for human-facing applications


Graham Neubig @gneubig · Jun 3

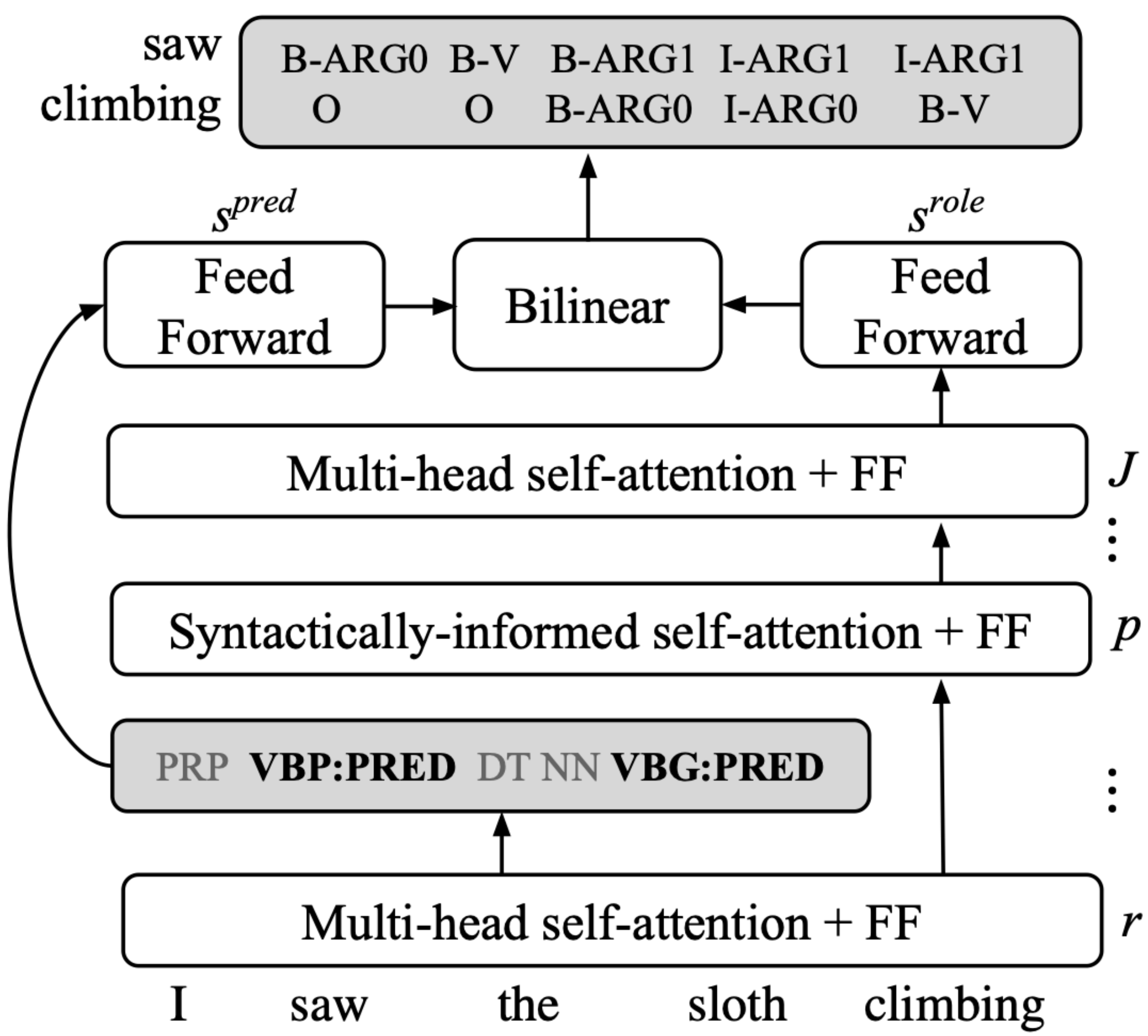
So @anas_ant and I were discussing "Is dependency parsing useful for anything in 2020?" It was more clear in 2010, but now most SOTA NLP models don't use dependencies as input. What are some really convincing use-cases of dependencies nowadays? The more the better!

39 35 150

https://twitter.com/gneubig/status/1268238606101032962?lang=en

Example 1: Adding Inductive Bias to Neural Models

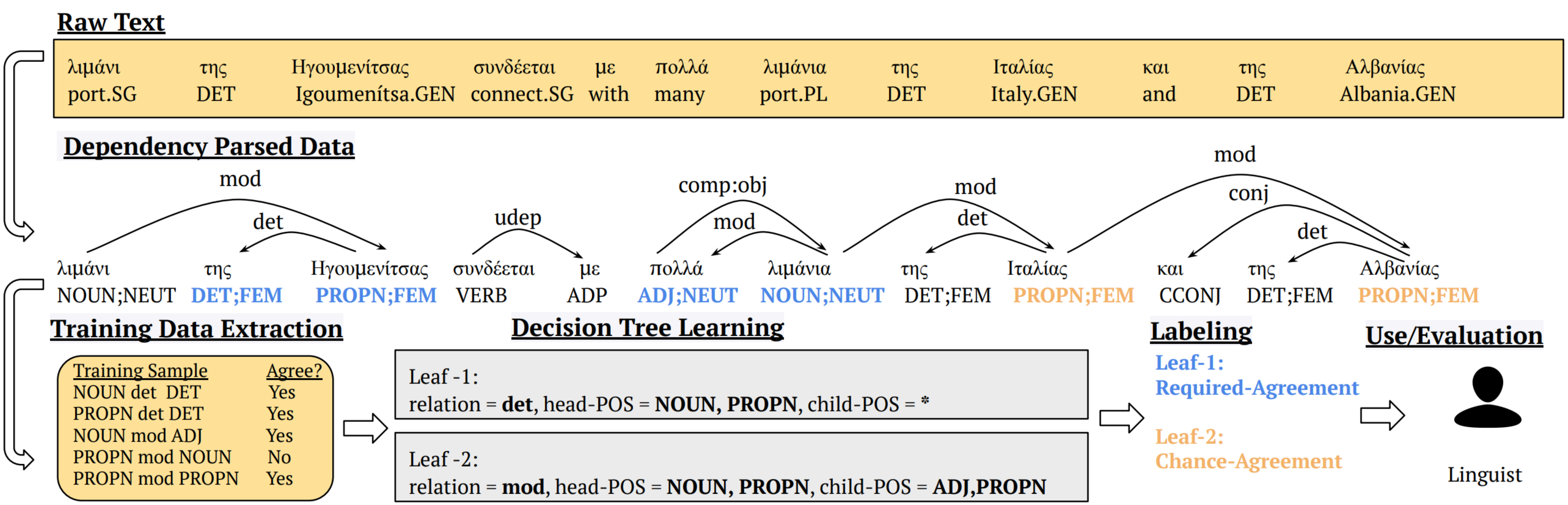
- Bias self attention to follow syntax



Strubell, Emma, et al. "Linguistically-informed self-attention for semantic role labeling." *arXiv preprint arXiv:1804.08199* (2018).

Example 2: Understanding Language Structure

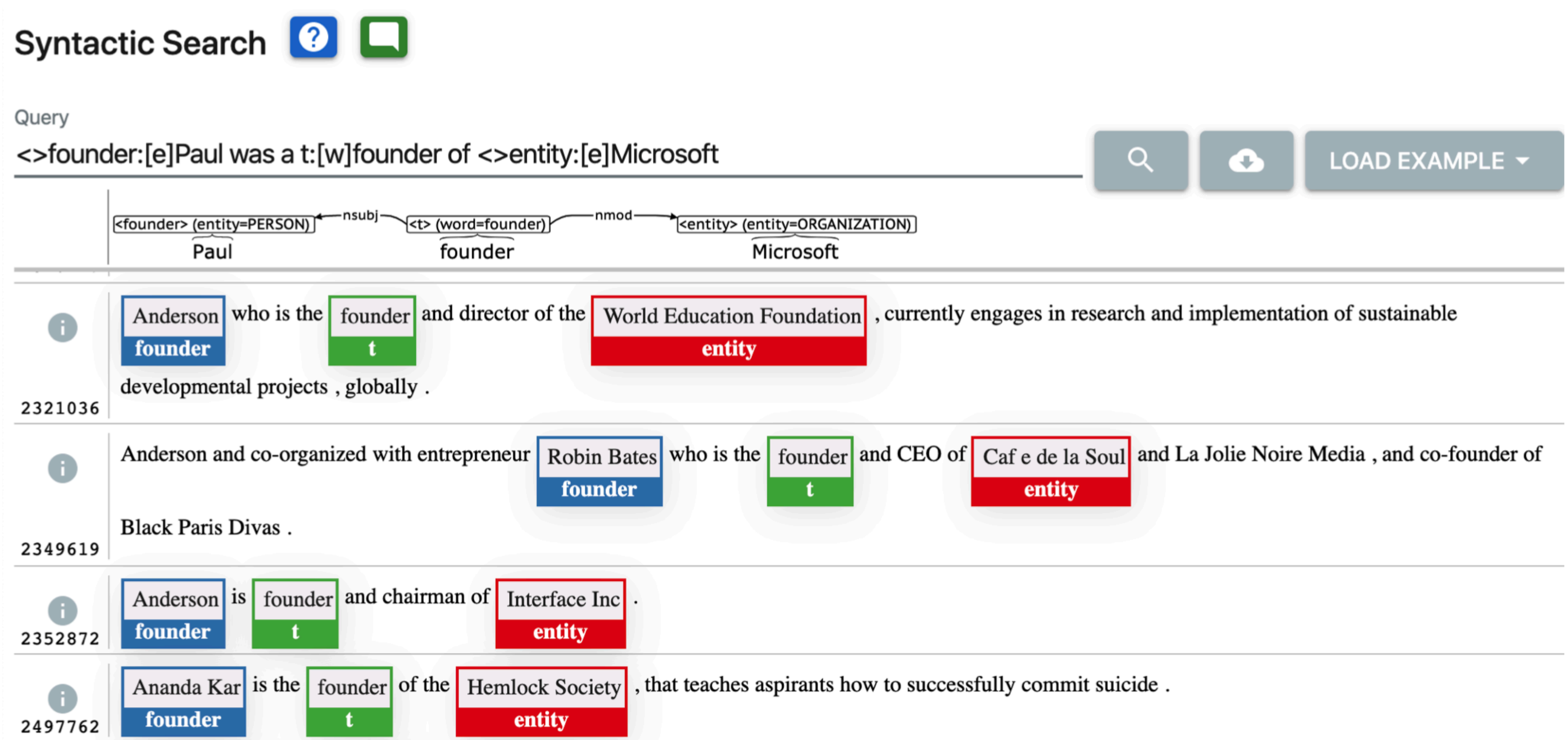
- Example of extracting morphological agreement rules using dependency relations



Chaudhary, Aditi, et al. "Automatic Extraction of Rules Governing Morphological Agreement." *EMNLP 2020*.

Example 3: Searching over Parsed Corpora

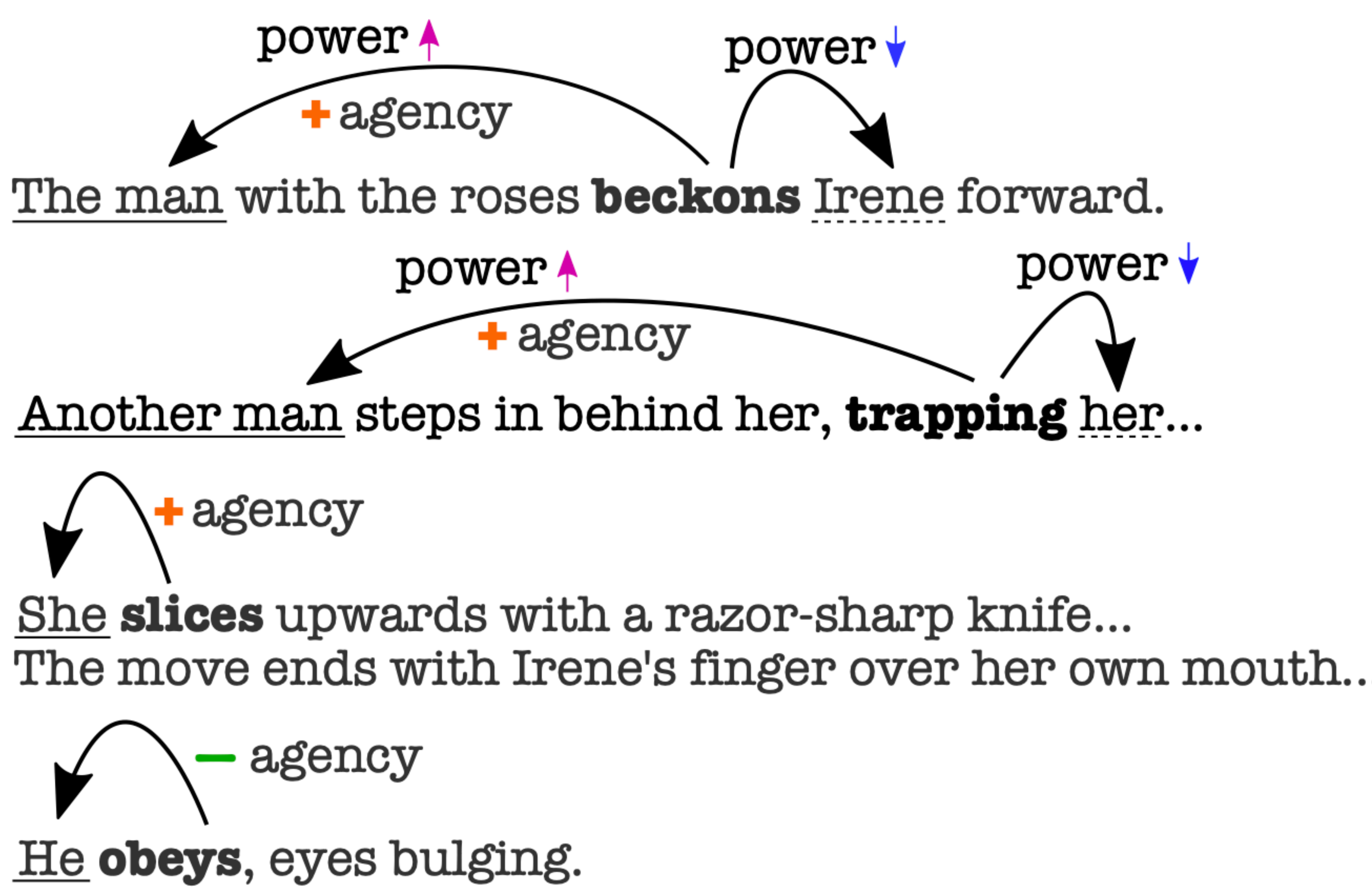
- Search using "syntactic regex"



Shlain, Micah, et al. "Syntactic Search by Example." *arXiv preprint arXiv:2006.03010* (2020).

Example 4: Analysis of Other Linguistic Phenomena

- Examining power and agency in film scripts



Sap, Maarten, et al. "Connotation frames of power and agency in modern films." *EMNLP 2017*.

Dependency Parsing

Parsing

- Predicting linguistic structure from input sentence
- **Transition-based models**
    - step through actions one-by-one until we have output
    - like history-based model for POS tagging
- **Graph-based models**
    - calculate probability of each edge/constituent, and perform some sort of dynamic programming
    - like linear CRF model for POS

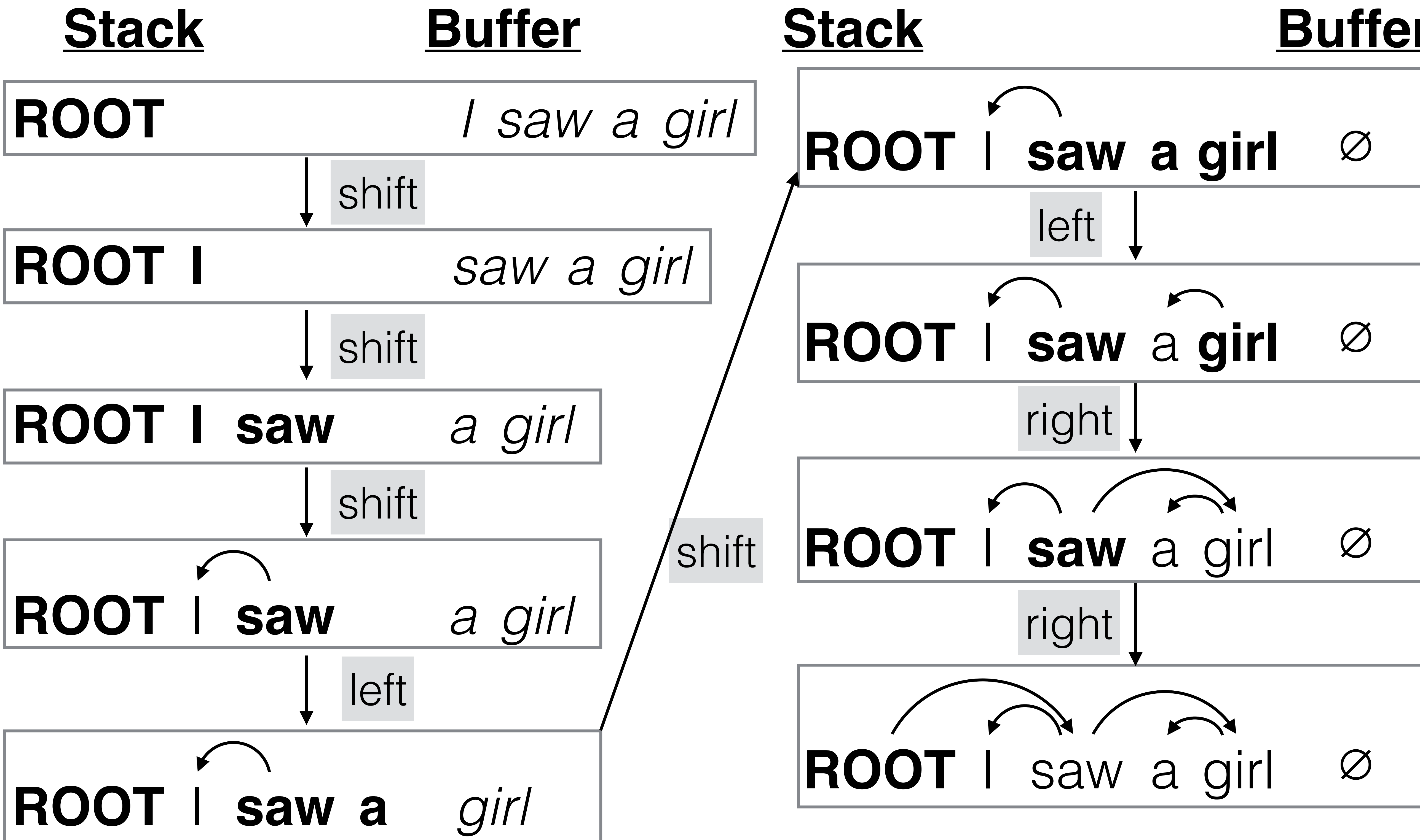
Shift-reduce Parsing

Arc Standard Shift-Reduce Parsing

(Yamada & Matsumoto 2003, Nivre 2003)

- Process words one-by-one left-to-right
- Two data structures
  - **Queue:** of unprocessed words
  - **Stack:** of partially processed words
- At each point choose
  - **shift:** move one word from queue to stack
  - **reduce left:** top word on stack is head of second word
  - **reduce right:** second word on stack is head of top word
- Learn how to choose each action with a classifier

Shift Reduce Example



Classification for Shift-reduce

- Given a **configuration**

Stack **Buffer**

ROOT | saw a *girl*

- Which **action** do we choose?

shift

ROOT | saw a girl ∅

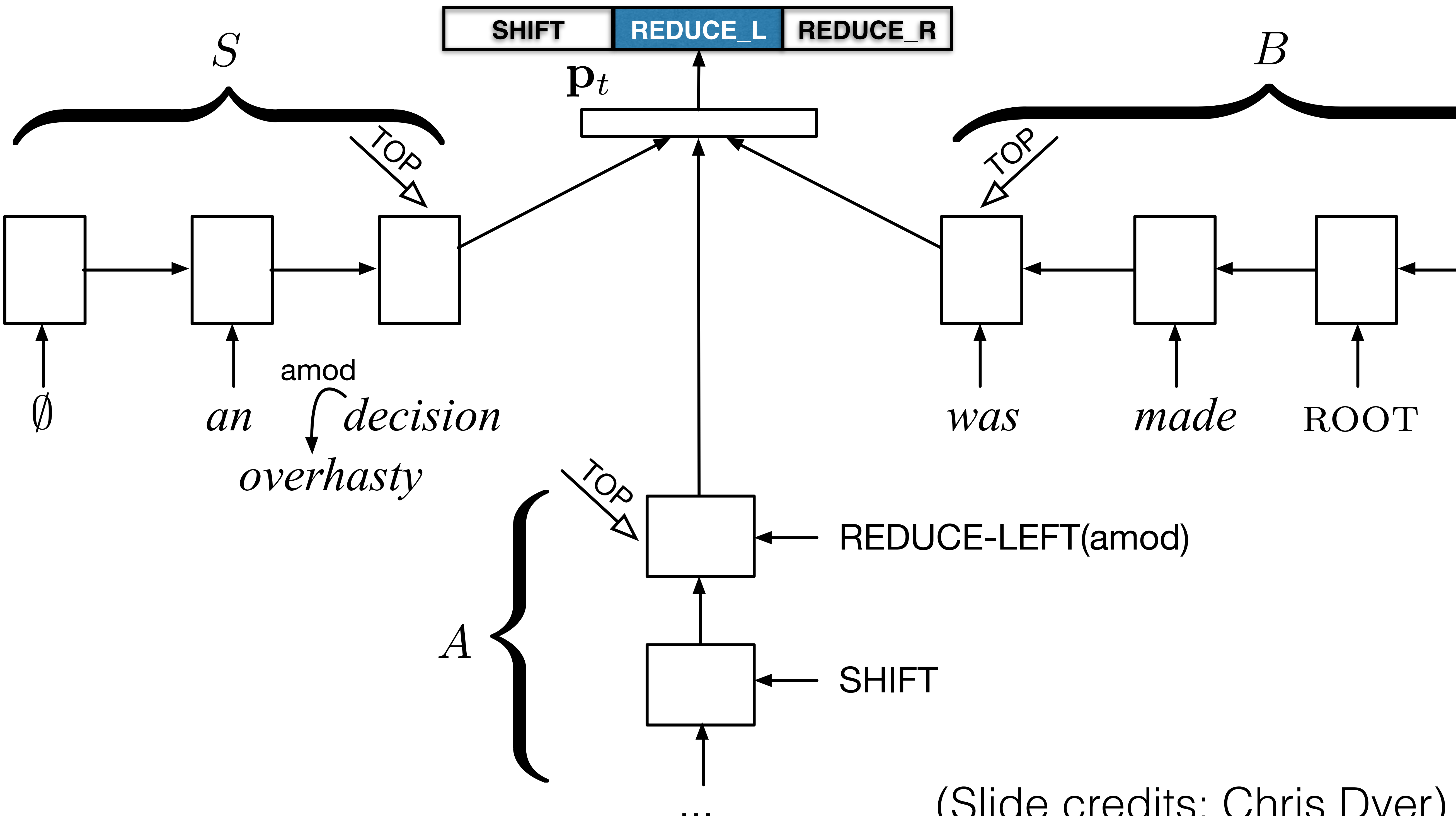
right

ROOT | saw a *girl*

left

ROOT | saw a *girl*

Encoding Stack Configurations w/ RNNs

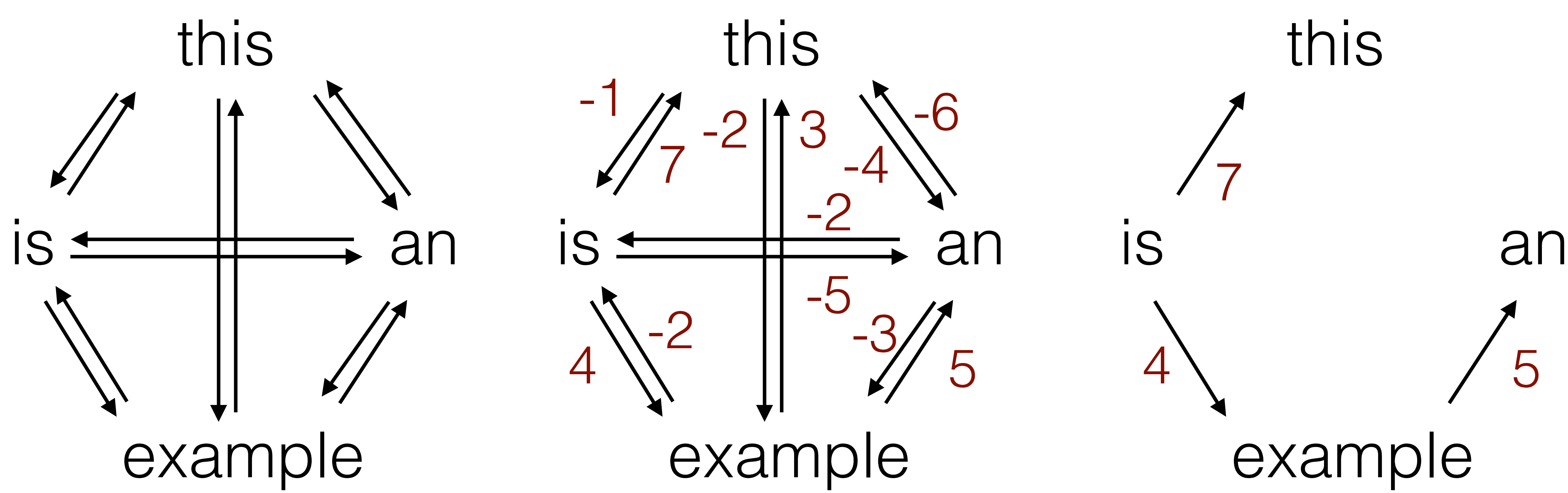


(Slide credits: Chris Dyer)

Graph-based Parsing

(First Order) Graph-based Dependency Parsing

- Express sentence as fully connected directed graph
- Score each edge independently
- Find maximal spanning tree



Graph-based vs. Transition Based

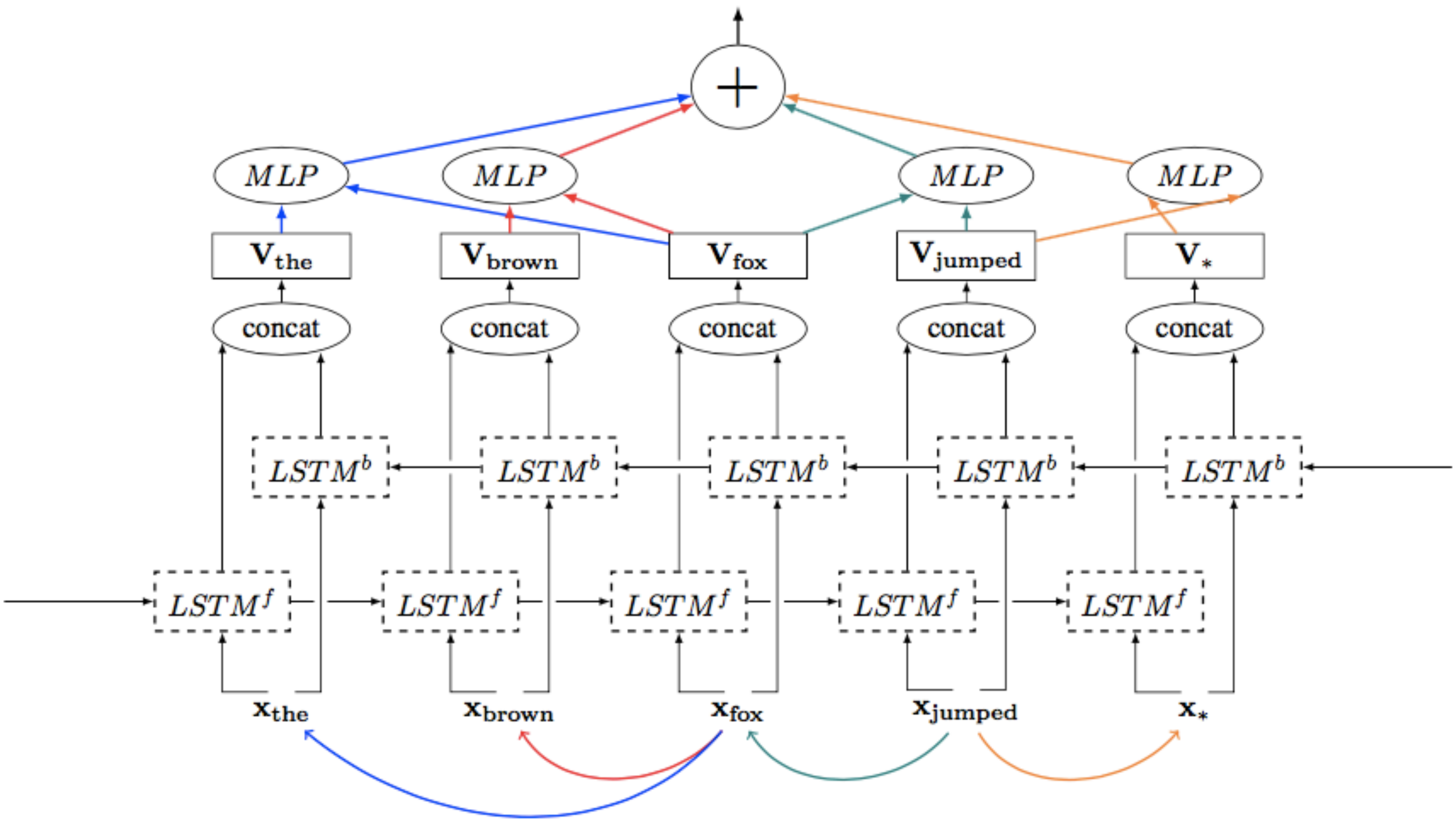
- **Transition-based**
  - + Easily condition on infinite tree context (structured prediction)
  - - Greedy search algorithm causes short-term mistakes
- **Graph-based**
  - + Can find exact best global solution via DP algorithm
  - - Have to make local independence assumptions

Chu-Liu-Edmonds (Chu and Liu 1965, Edmonds 1967)

- We have a graph and want to find its spanning tree
- **Greedily select** the best incoming edge to each node (and subtract its score from all incoming edges)
- If there are cycles, select a cycle and **contract** it into a single node
- **Recursively call** the algorithm on the graph with the contracted node
- **Expand** the contracted node, deleting an edge appropriately

Sequence Model Feature Extractors

(Kipperwasser and Goldberg 2016)



BiAffine Classifier
(Dozat and Manning 2017)

$$\mathbf{h}_i^{(arc\text{-}dep)} = \mathbf{MLP}^{(arc\text{-}dep)}(\mathbf{r}_i)$$
$$\mathbf{h}_j^{(arc\text{-}head)} = \mathbf{MLP}^{(arc\text{-}head)}(\mathbf{r}_j)$$
$$\mathbf{s}_i^{(arc)} = H^{(arc\text{-}head)} U^{(1)} \mathbf{h}_i^{(arc\text{-}dep)} + H^{(arc\text{-}head)} \mathbf{u}^{(2)}$$

Learn specific representations for head/dependent for each word

Calculate score of each arc

- Just optimize the likelihood of the parent, no structured training
  - This is a local model, with global decoding using MST at the end
- Best results (with careful parameter tuning) on universal dependencies parsing task

Multilingual Dependency Parsing

Difficulty In Multilingual Dependency Parsing

- Syntactic analysis is a *particularly hard* multilingual task
- It is on the *global level*, not just word-by-word level
- Syntax *varies widely* across different languages

Example Improvement 1: Order-insensitive Encoders

- Standard cross-lingual transfer can fail with large word order differences between source and target
- Change model structure to be order-insensitive to avoid over-fitting to source

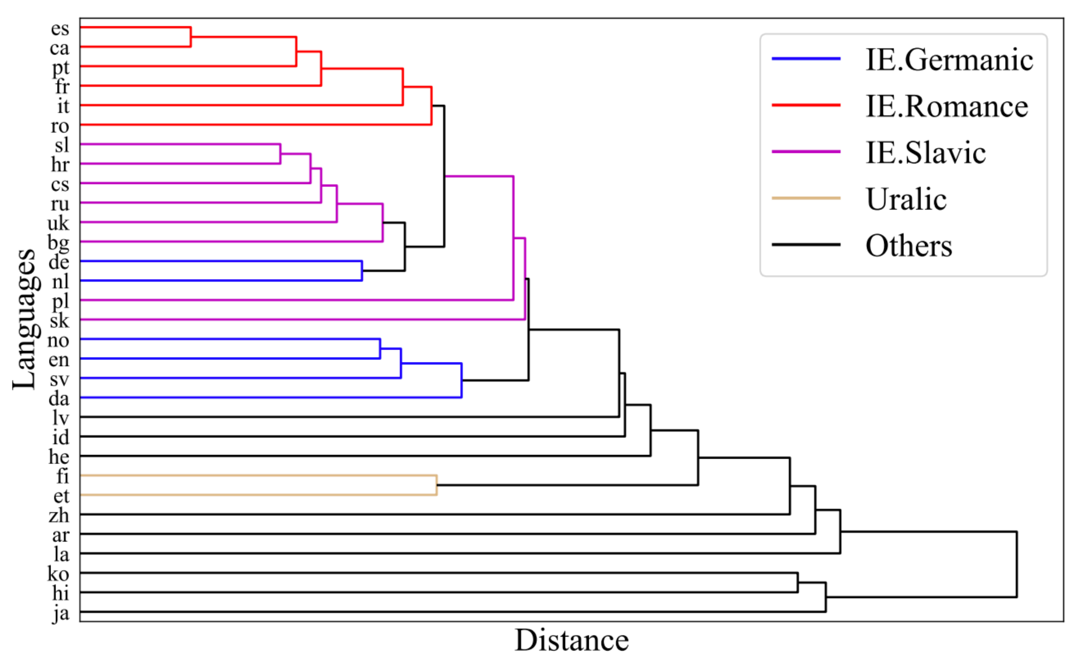


Figure 1: Hierarchical clustering (with the Nearest Point Algorithm) dendrogram of the languages by their word-ordering vectors.

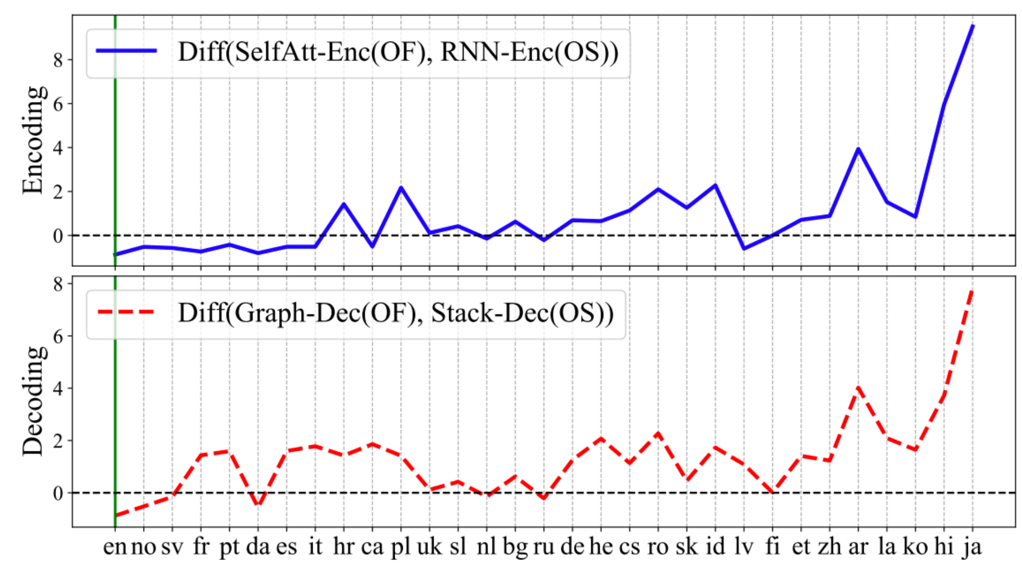
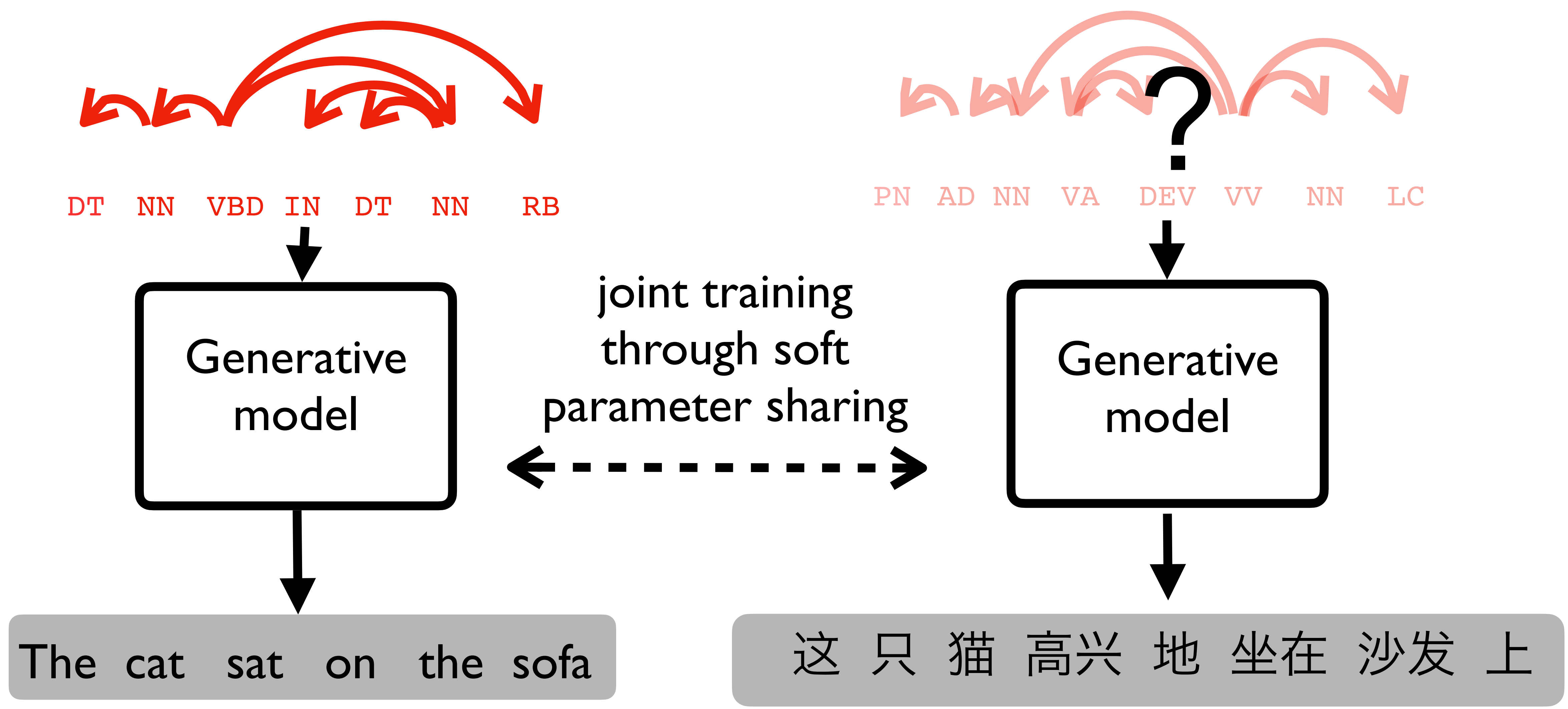


Figure 2: Evaluation score differences between Order-Free (OF) and Order Sensitive (OS) modules. We show results of both encoder (blue solid curve) and decoder (dashed red curve). Languages are sorted by their word-ordering distances to English from left to right. The position of English is marked with a green bar.

Ahmad, Wasi Uddin, et al. "On difficulties of cross-lingual transfer with order differences: A case study on dependency parsing." *NAACL 2019*.

Example Improvement 2: Generative Model Fine-tuning

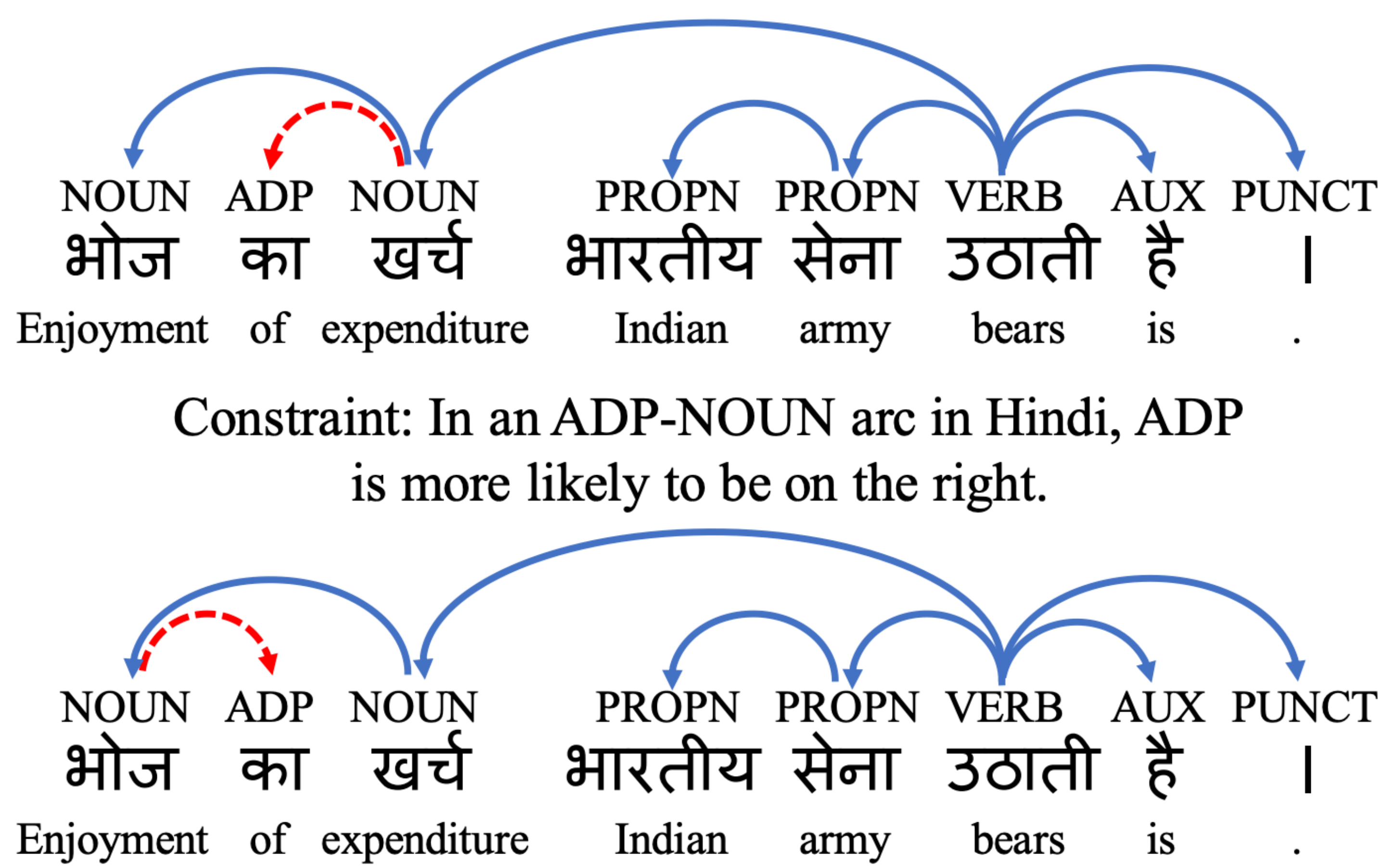
- Use *generative model* that can be trained unsupervised, and fine-tune on the target language



He, Junxian, et al. "Cross-lingual syntactic transfer through unsupervised adaptation of invertible projections." *ACL 2019*.

Example Improvement 3: Linguistically Informed Constraints

- Add constraints based on a-priori knowledge of the language structure



Meng, Tao, Nanyun Peng, and Kai-Wei Chang. "Target language-aware constrained inference for cross-lingual dependency parsing." *EMNLP 2019*.

Discussion Question

Discussion Question

- Read at least one of the three below papers on cross-lingual dependency parsing

- What do you think went well, and what ideas do you have to further improve?

Ahmad, Wasi Uddin, et al. "On difficulties of cross-lingual transfer with order differences: A case study on dependency parsing." *NAACL 2019*.

He, Junxian, et al. "Cross-lingual syntactic transfer through unsupervised adaptation of invertible projections." *ACL 2019*.

Meng, Tao, Nanyun Peng, and Kai-Wei Chang. "Target language-aware constrained inference for cross-lingual dependency parsing." *EMNLP 2019*.