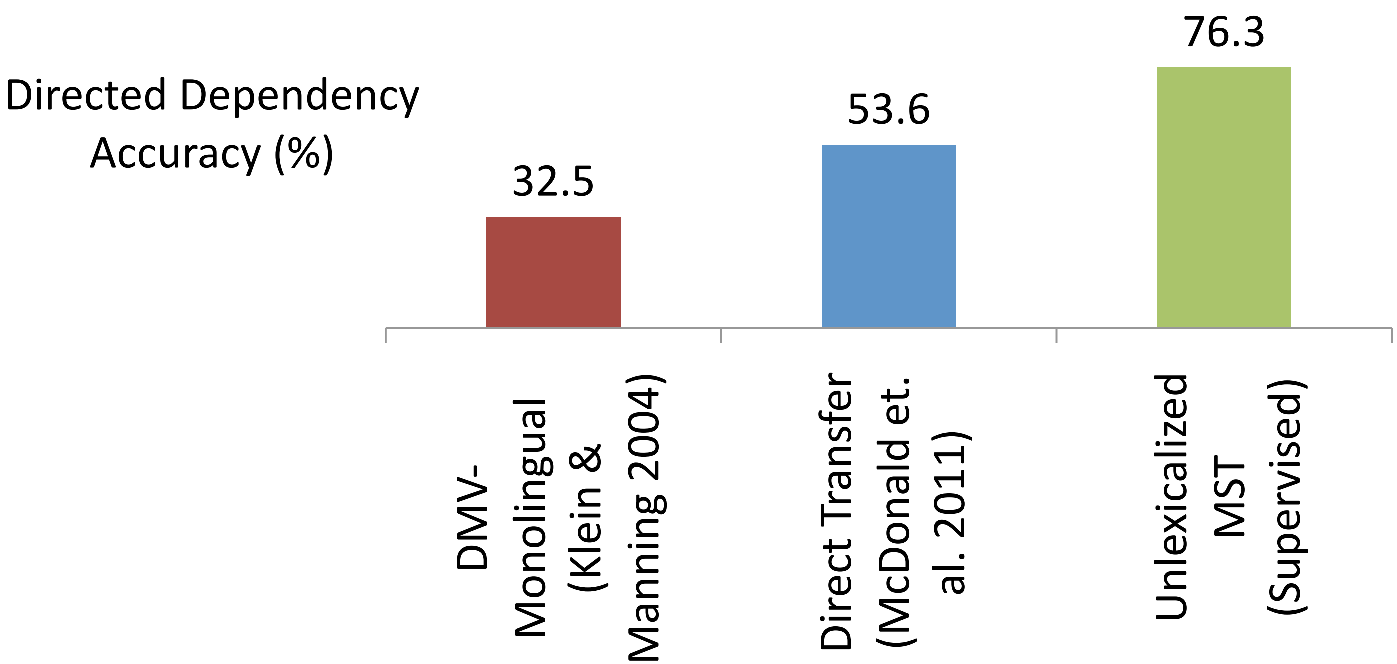
Selective-Sharing for Multilingual Dependency Parsing

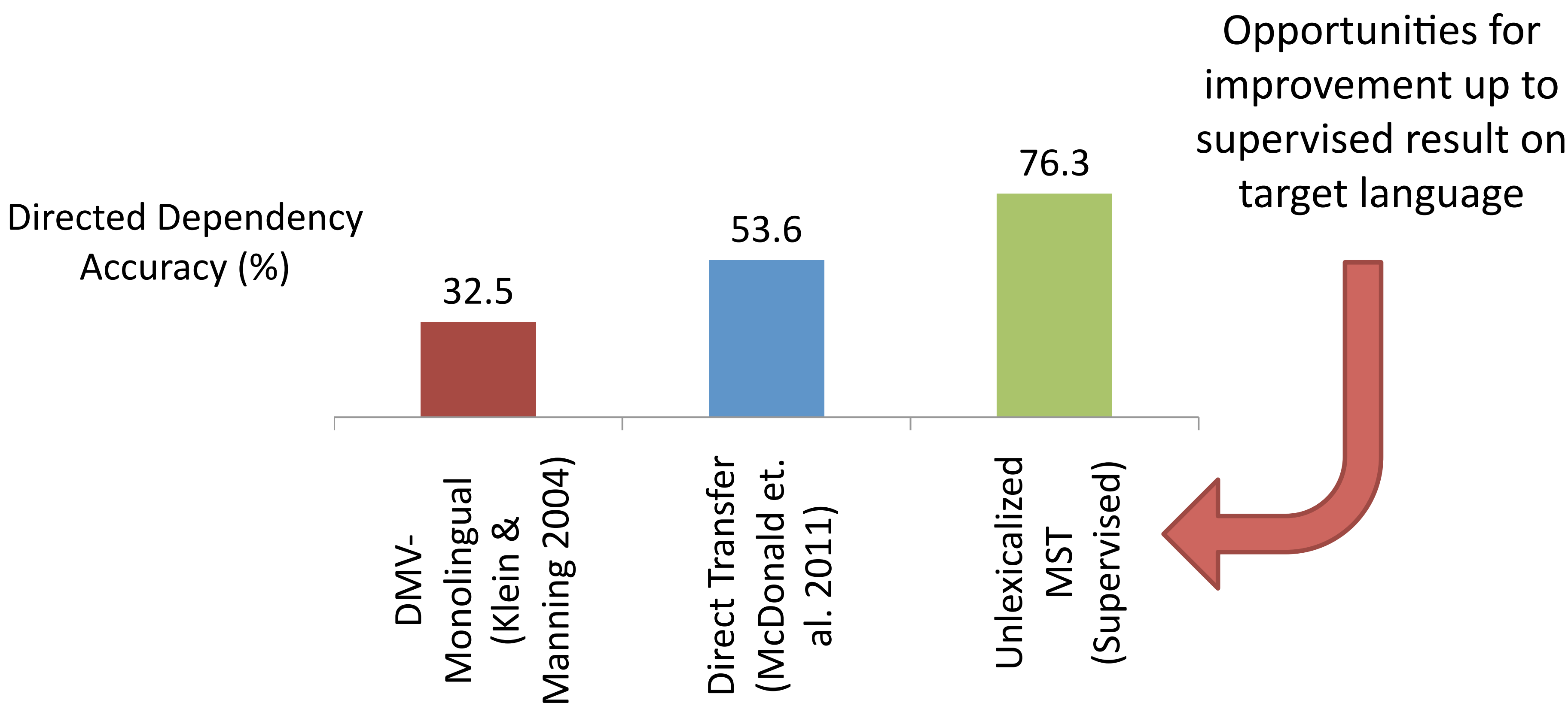
Zach Hynes[1]
Tahira Naseem[1], Regina Barzilay[1], Amir Globerson[2]
MIT, CSAIL[1]
Hebrew University[2]

- Multilingual parsing is essential for text analysis in resource-poor languages (e.g. tree-to-tree MT systems)
- Existing methods do not capture linguistic similarities and differences



Results averaged over 17 languages

- Multilingual parsing is essential for text analysis in resource-poor languages (e.g. tree-to-tree MT systems)
- Existing methods do not capture linguistic similarities and differences



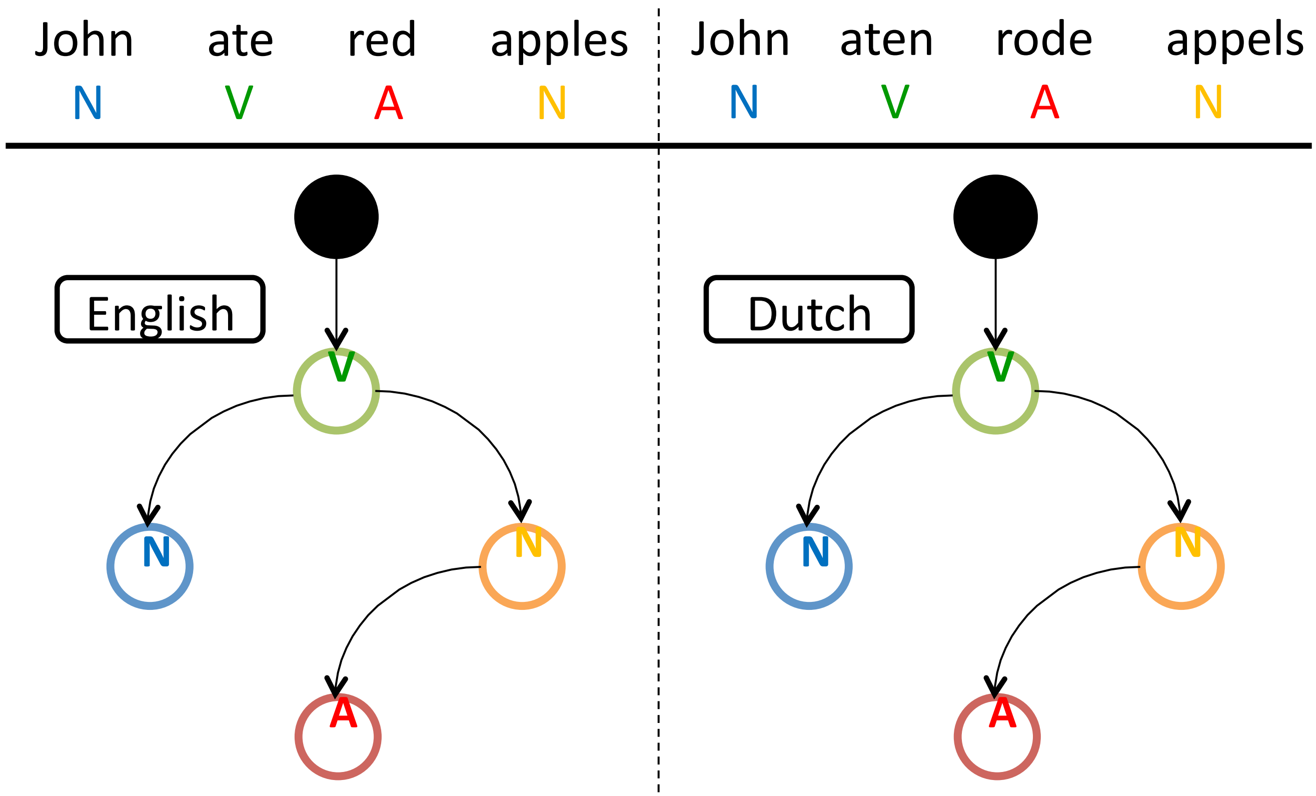
Results averaged over 17 languages

Crosslingual Parser Transfer

$\boxed{\text{Motivated by Syntactic Similarities across Languages}}$

Crosslingual Parser Transfer

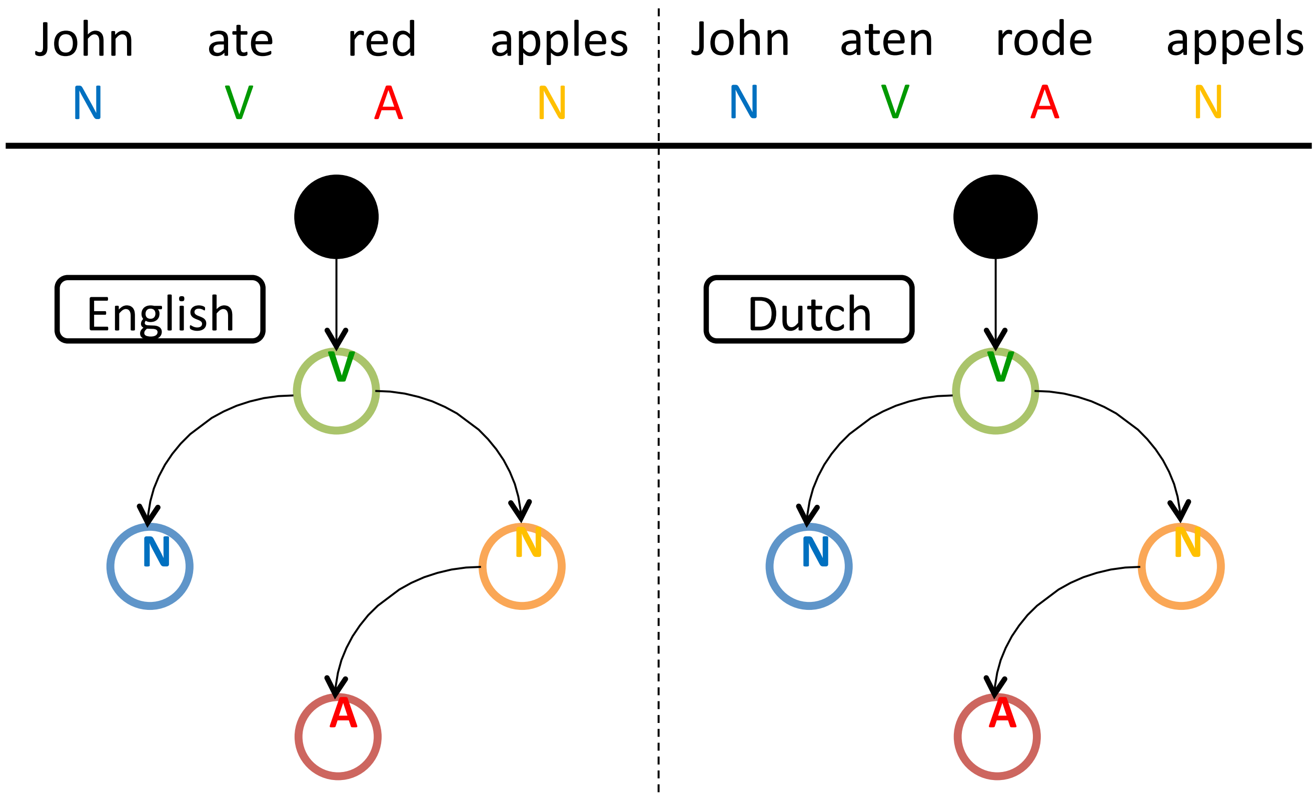
Motivated by Syntactic Similarities across Languages



Crosslingual Parser Transfer

Motivated by Syntactic Similarities across Languages

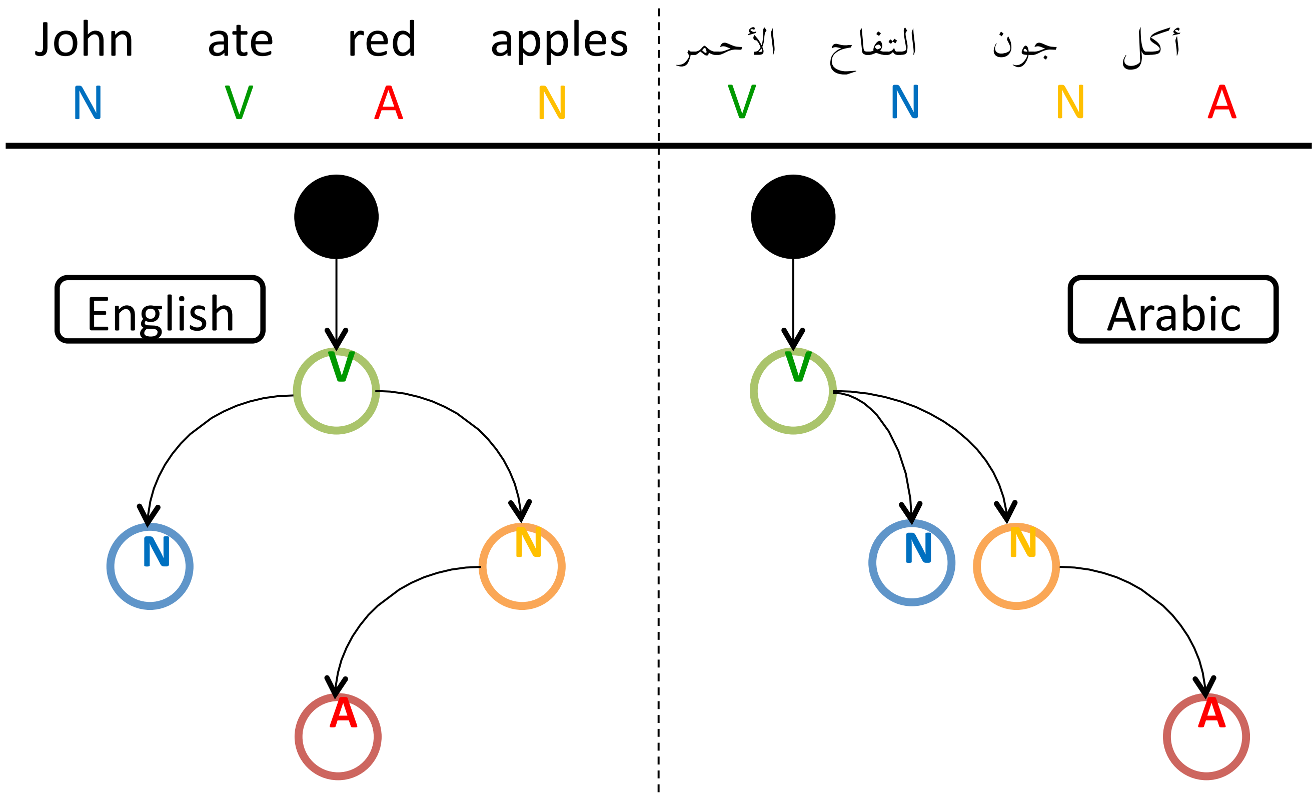
Q. What if the target language is very different?



Crosslingual Parser Transfer

Motivated by Syntactic Similarities across Languages

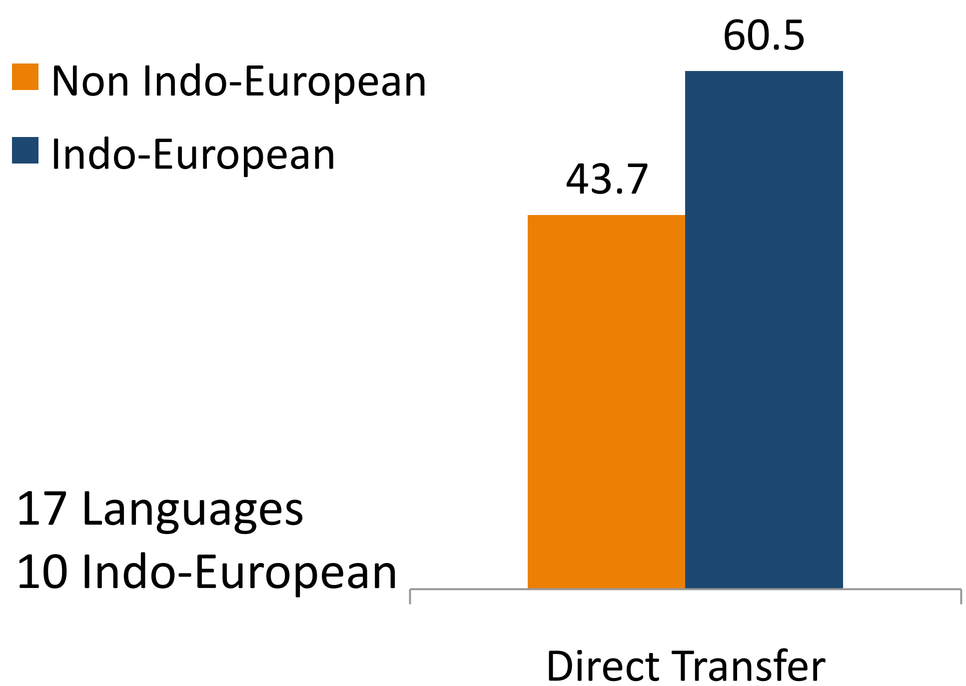
Q. What if the target language is very different?



Crosslingual Parser Transfer

Motivated by Syntactic Similarities across Languages

Q. What if the target language is very different?
A. Existing transfer approaches are not effective.



Crosslingual Parser Transfer

Motivated by Syntactic Similarities across Languages

Q. What if the target language is very different?
A. Existing transfer approaches are not effective.

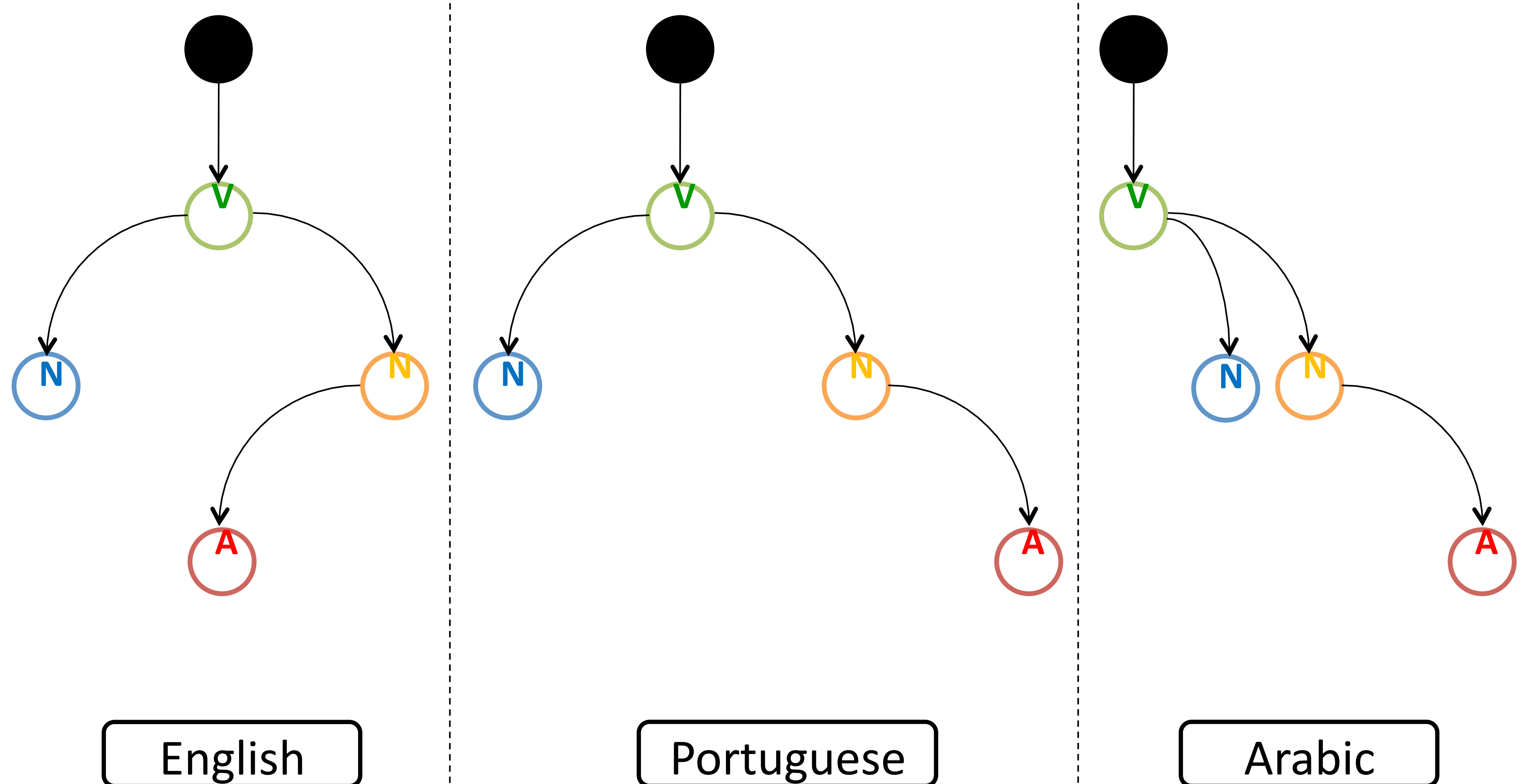
Our approach: Selective Sharing

Accounts for both **Similarities** and **Differences**

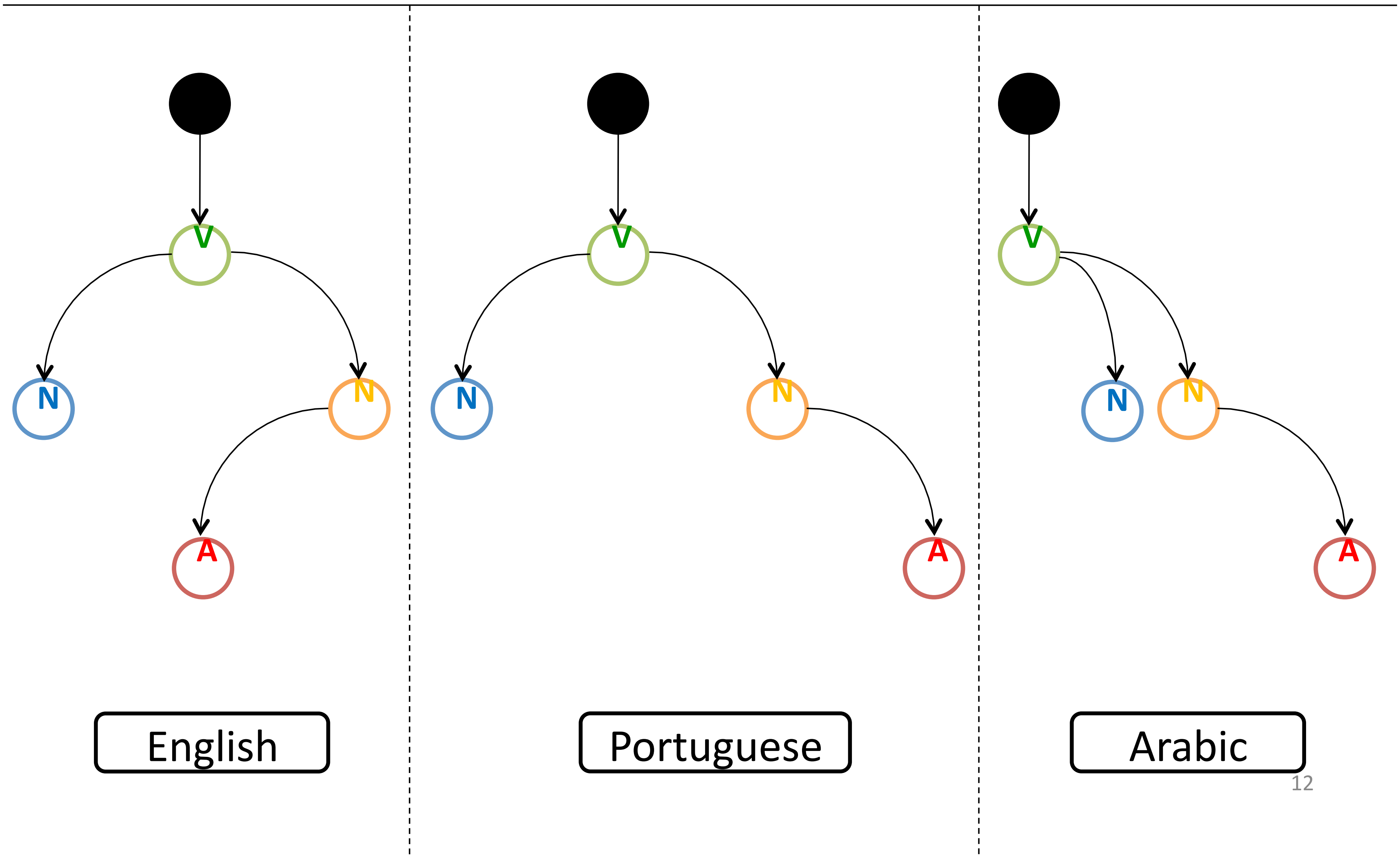
Selective Sharing: Linguistic Basis

- **Linguistic theories:** a unified view of syntax that encodes crosslingual **similarities**
  - The definition of POS tags (Nouns, Verbs etc.) is the same across languages
  - All languages are analyzed into NP's, VP's, PP's etc.
- **Linguistic typology:** systematic study of crosslingual **differences**
  - Prefixing vs. Postfixing languages
  - Preposition vs. Postposition

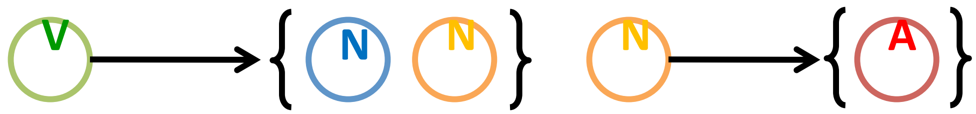
John	ate	red	apples	João	comeu	maçãs	vermelhas	الأحمر	التفاح	جون	أكلت
N	V	A	N	N	V	N	A	V	N	N	A

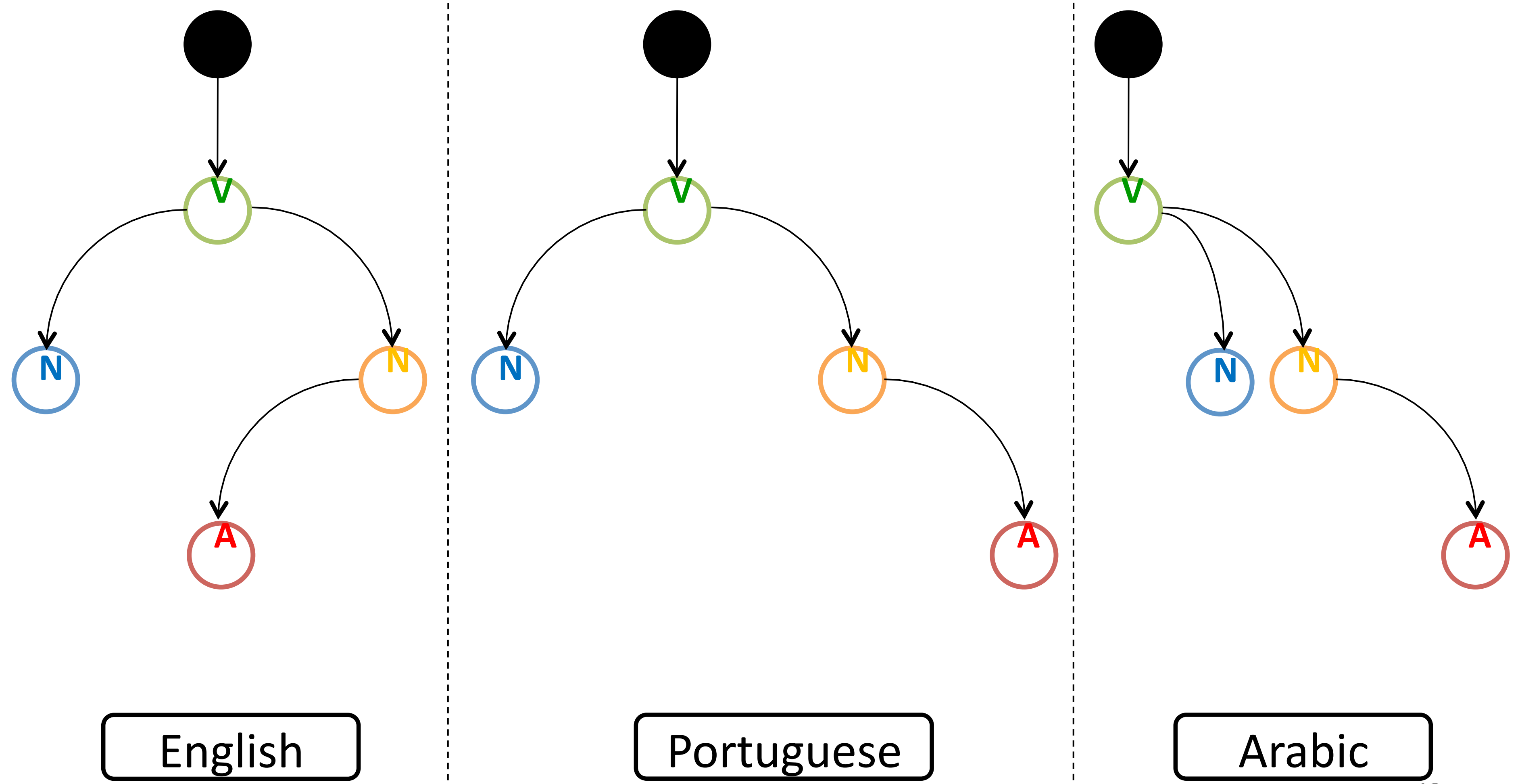


What is language-universal?

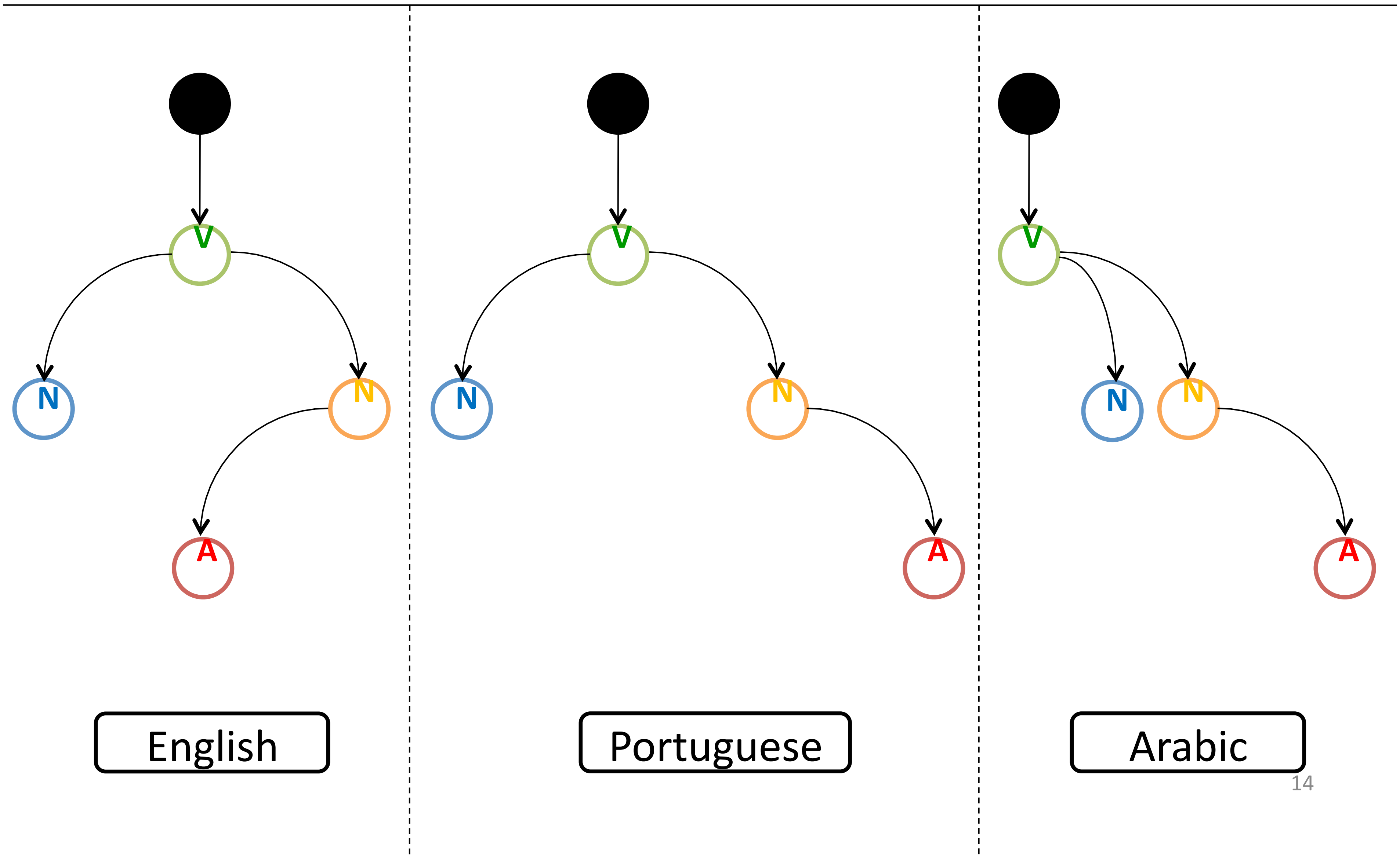


What is language-universal? **dependent "*Selection*"**

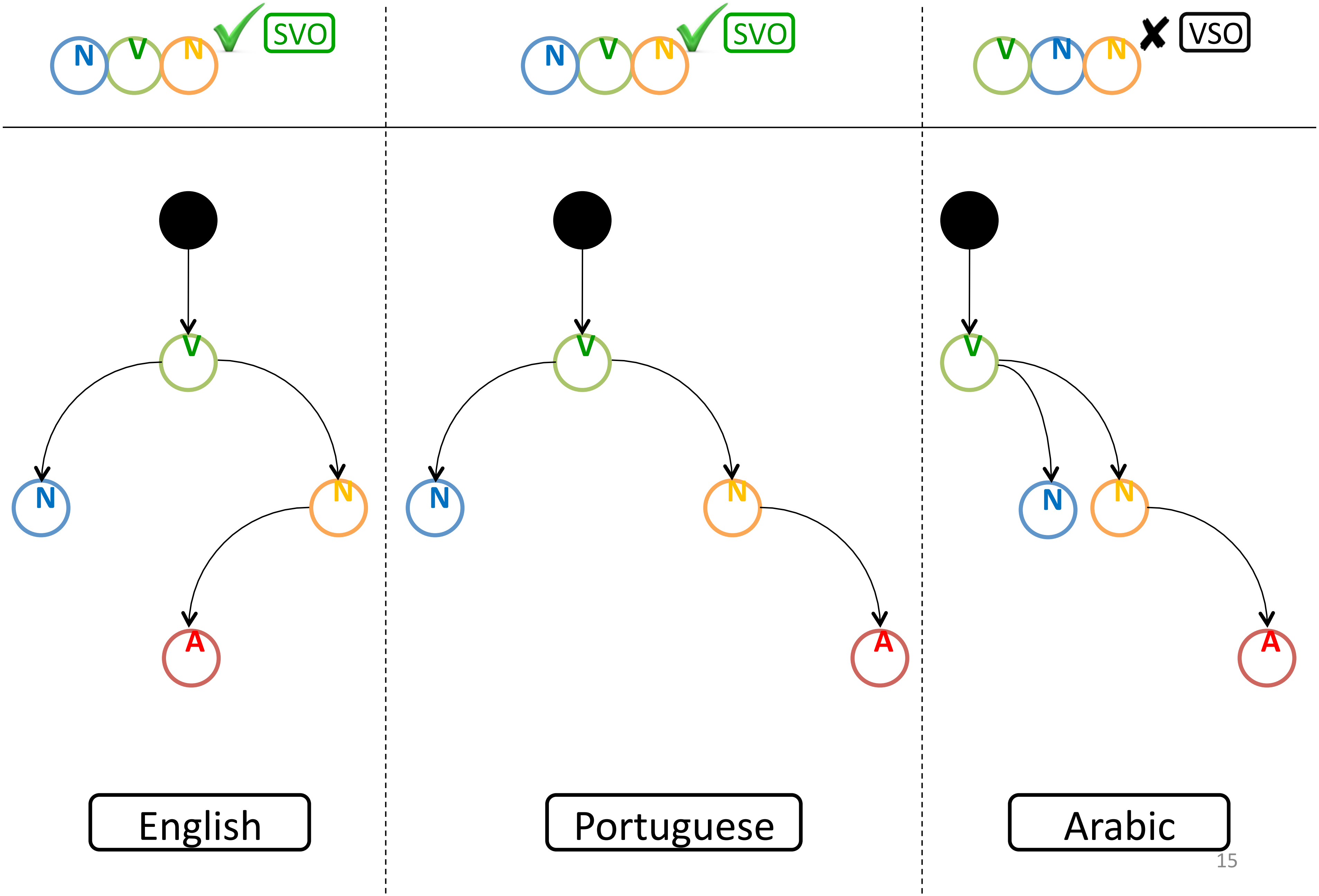




What is language-specific? … **ordering**



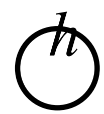
"Ordering" typological features can guide



Model

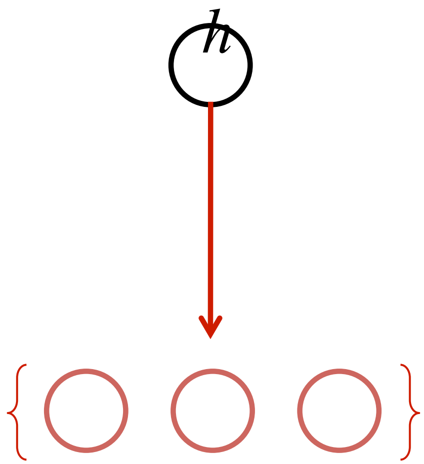
- Separate "*Selection*" from "*Ordering*"
  - *Selection* is fully shared across all languages
  - *Ordering* is selectively shared across languages
    (sharing is guided by typology)

Generative Process



Generative Process: Selection

$n = 3$



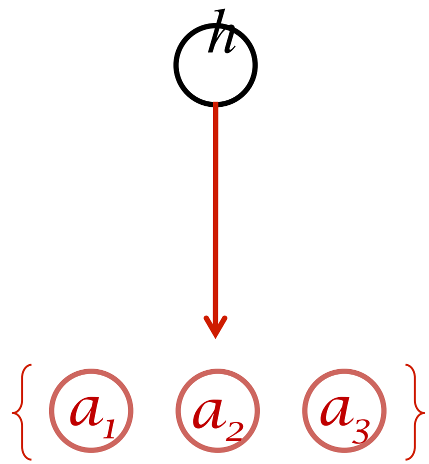
Select the number n of arguments

$$P(n|h) = \theta(n|h)$$

Generative Process: Selection

$n = 3$

$S = \{a_1, a_2, a_3\}$



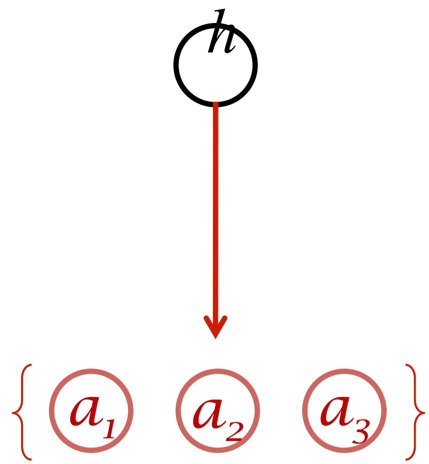
Select the set S of arguments

$$P(S|h,n) = \frac{e^{\mathbf{w}_{sel} \cdot \mathbf{f}(S,h,n)}}{Z_{h,n}}$$

Generative Process: Selection

$n = 3$

$S = \{a_1, a_2, a_3\}$



Only pair-wise parent-child features

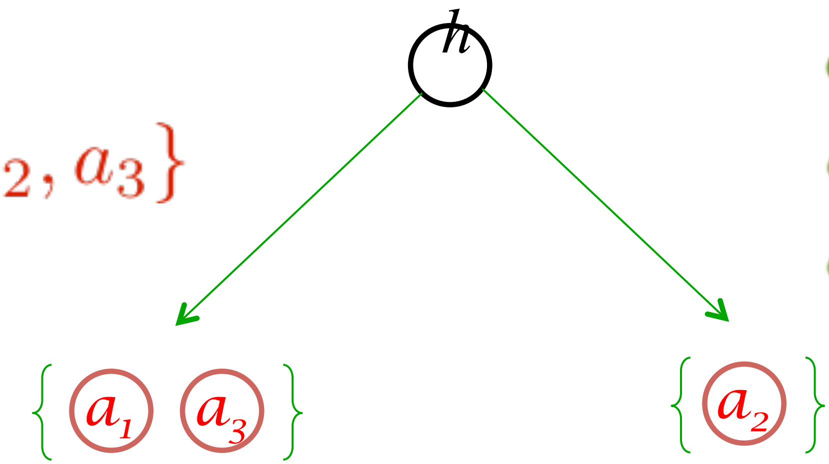
Select the set S of arguments

$$P(S|h,n) = \frac{e^{\mathbf{w}_{sel} \cdot \mathbf{f}(S,h,n)}}{Z_{h,n}}$$

Generative Process: Ordering

$n = 3$

$S = \{a_1, a_2, a_3\}$



$dir^{(a_1)} = L$

$dir^{(a_3)} = L$

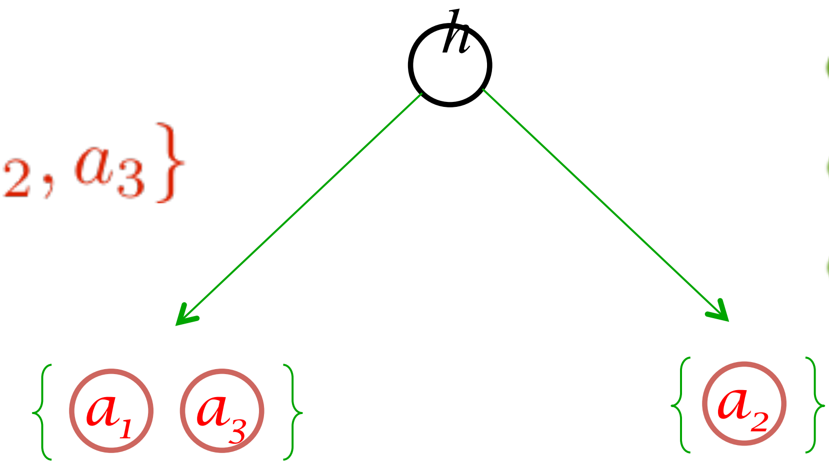
$dir^{(a_2)} = R$

for each argument select dir (right/left) w.r.t. head

$$P(dir|a, h, l) = \frac{e^{\mathbf{w}_{ord} \cdot \mathbf{f}(dir, a, h, \mathbf{v}_l)}}{Z_{a,h,l}}$$

Generative Process: Ordering

$$n = 3$$
$$S = \{a_1, a_2, a_3\}$$



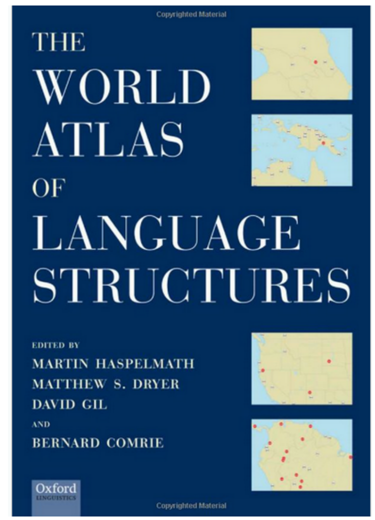
$$dir(a_1) = L$$
$$dir(a_3) = L$$
$$dir(a_2) = R$$

Ordering depends on language specific features

$$P(dir|a, h, l) = \frac{e^{\mathbf{w}_{ord} \cdot \mathbf{f}(dir, a, h, \mathbf{v}_l)}}{Z_{a,h,l}}$$

Generative Process: Ordering Guided by Typology

Typological Features
1. Order of Subject, Object and Verb
2. Order of Adposition and Noun
3. Order of Genitive and Noun
4. Order of Adjective and Noun
5. Order of Demonstrative and Noun
6. Order of Numeral and Noun

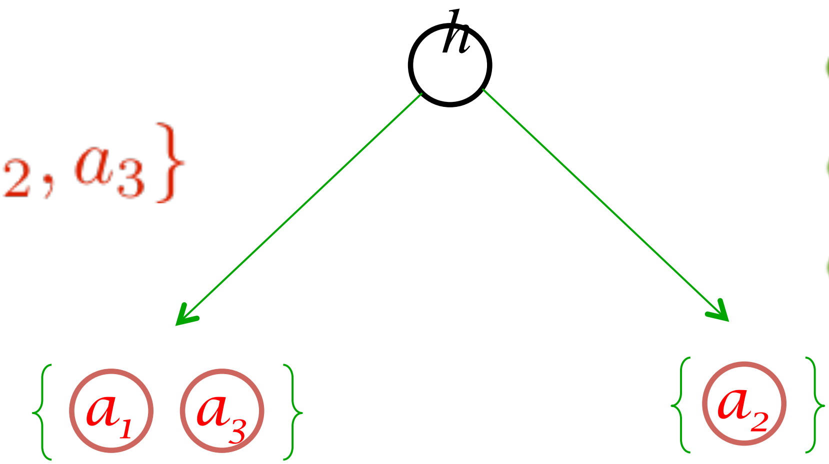


- ➢Correspondence between the features and the parameters is learned automatically.
- ➢These features do not cover all ordering decisions.

Generative Process: Ordering

$n = 3$

$S = \{a_1, a_2, a_3\}$



$dir(a_1) = L$

$dir(a_3) = L$

$dir(a_2) = R$

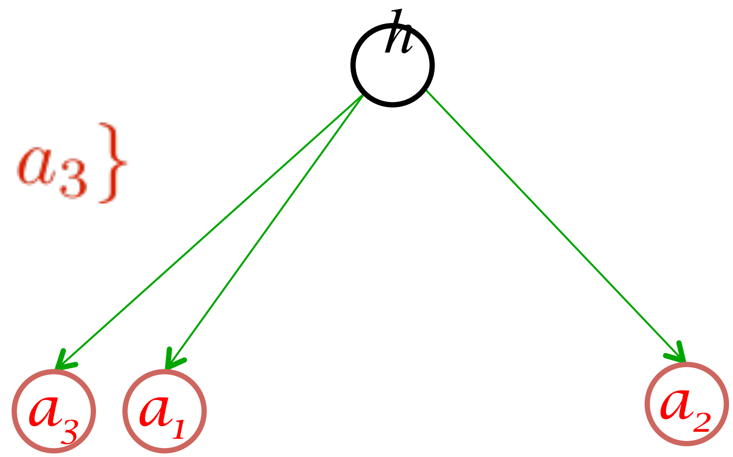
Ordering depends on language specific features

$$P(dir|a, h, l) = \frac{e^{\mathbf{w}_{ord} \cdot \mathbf{f}(dir, a, h, \mathbf{v}_l)}}{Z_{a,h,l}}$$

Generative Process: Ordering

$n = 3$

$S = \{a_1, a_2, a_3\}$



$dir(a_1) = L$

$dir(a_3) = L$

$dir(a_2) = R$

Select the internal order of right/left arguments

Uniformly from among all possible orderings

Model Parameters

- Distribution over number of arguments given the parent tag $\theta(n|h)$
- Weights for selection features, shared across all set sizes

$$P(S|h,n) = \frac{e^{\mathbf{w}_{sel} \cdot \mathbf{f}(S,h,n)}}{Z_{h,n}}$$

- Weights for ordering features

$$P(dir|a,h,l) = \frac{e^{\mathbf{w}_{ord} \cdot \mathbf{f}(dir,a,h,\mathbf{v}_l)}}{Z_{a,h,l}}$$

Model Parameters

- Distribution over number of arguments given the parent tag $\theta(n|h)$
- Weights for selection features, shared across all set sizes

$$P(S|h,n) = \frac{e^{\mathbf{w}_{sel}\cdot\mathbf{f}(S,h,n)}}{Z_{h,n}}$$

- Weights for ordering features

$$P(dir|a,h,l) = \frac{e^{\mathbf{w}_{ord}\cdot\mathbf{f}(dir,a,h,\mathbf{v}_l)}}{Z_{a,h,l}}$$

All parameters are shared across languages

Model Parameters

- Distribution over number of arguments given the parent tag $\theta(n|h)$
- Weights for selection features, shared across all set sizes

$$P(S|h,n) = \frac{e^{\mathbf{w}_{sel}\cdot\mathbf{f}(S,h,n)}}{Z_{h,n}}$$

- Weights for ordering features

$$P(dir|a,h,l) = \frac{e^{\mathbf{w}_{ord}\cdot\mathbf{f}(dir,a,h,\mathbf{v}_l)}}{Z_{a,h,l}}$$

All parameters are shared across languages

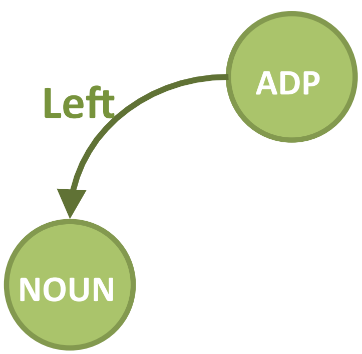
Setup

- For all languages:
  - Observed part-of-speech tags
  - Ordering features
    - **Either** observed typology for all languages
    - **Or** latent binary features for all languages
- For helper languages:
  - Observed dependency trees
- For unsupervised language:
  - Latent dependency trees

Analysis of Ordering Feature Weights

Typological Features

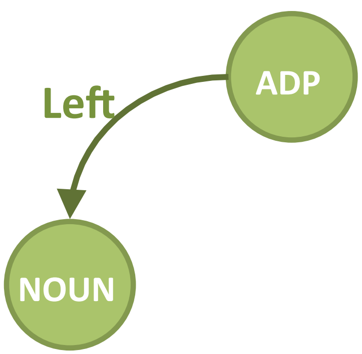
f (L, NOUN, ADP,)



f
L+NOUN+ADP+Postposition
L+NOUN+ADP+SOV
.....
.....
.....
L+NOUN+ADP+SVO
L+NOUN+ADP+Prepositions

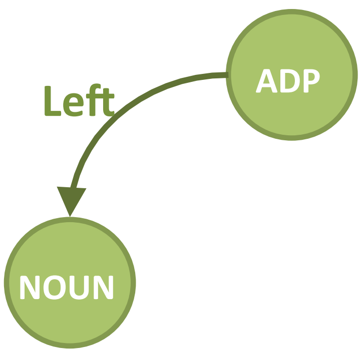
Analysis of Ordering Feature Weights

Highest Weight



f	w
L+NOUN+ADP+Postposition	1.6
L+NOUN+ADP+SOV	1.1
.....	.....
.....	.....
.....	.....
L+NOUN+ADP+SVO	-0.7
L+NOUN+ADP+Prepositions	-1.5

Analysis of Ordering Feature Weights



f	w
L+NOUN+ADP+Postposition	1.6
L+NOUN+ADP+SOV	1.1
.....	.....
.....	.....
.....	.....
L+NOUN+ADP+SVO	-0.7
L+NOUN+ADP+Prepositions	-1.5

Lowest Weight

Learning

- Optimizes the log-likelihood of data
  - Standard Expectation Maximization
    - Marginalize over all latent variables
    - M-Step requires gradient search
  - Posterior Constraint on dependency length

(Graca et. al. 2008)

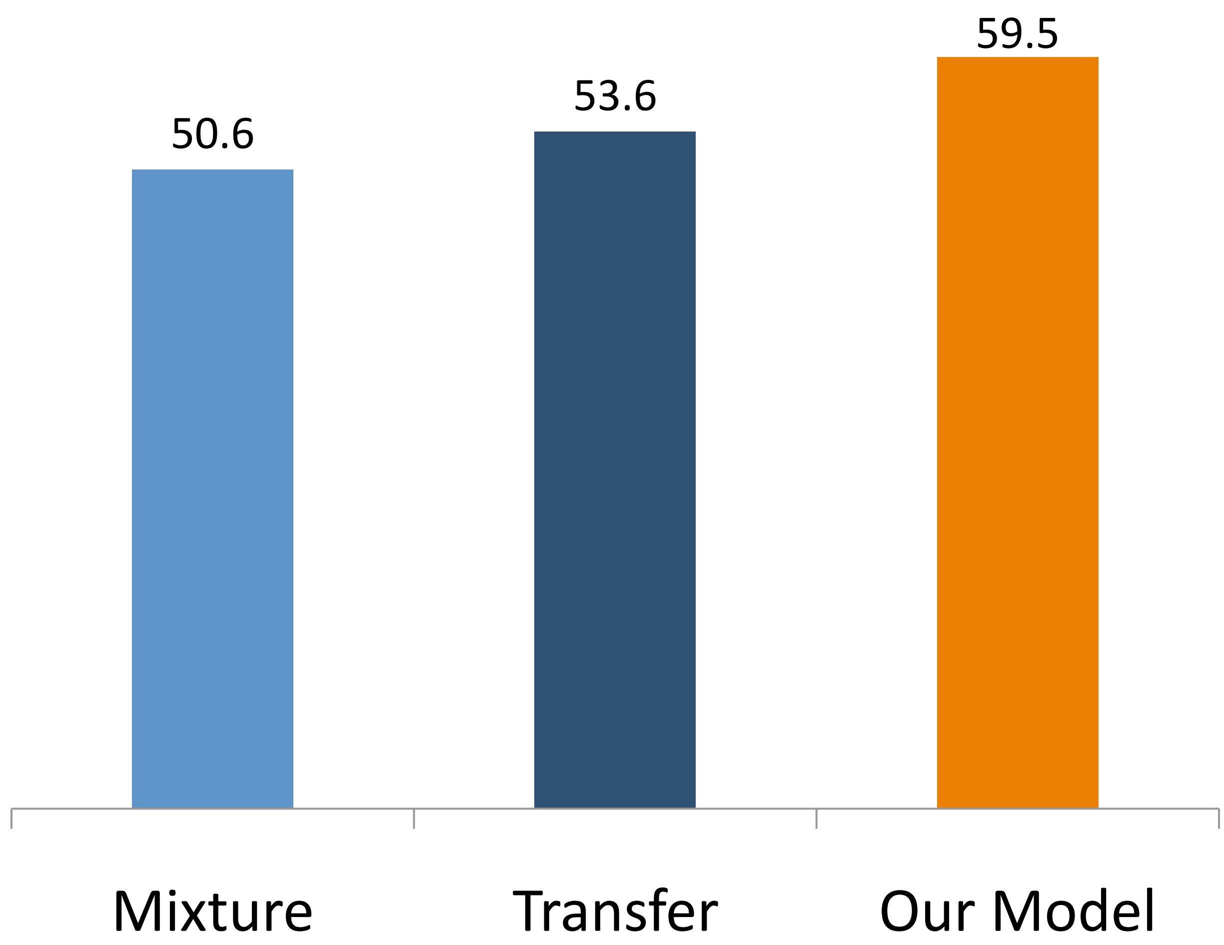
Experiments & Results

- **17** languages from **10** families:

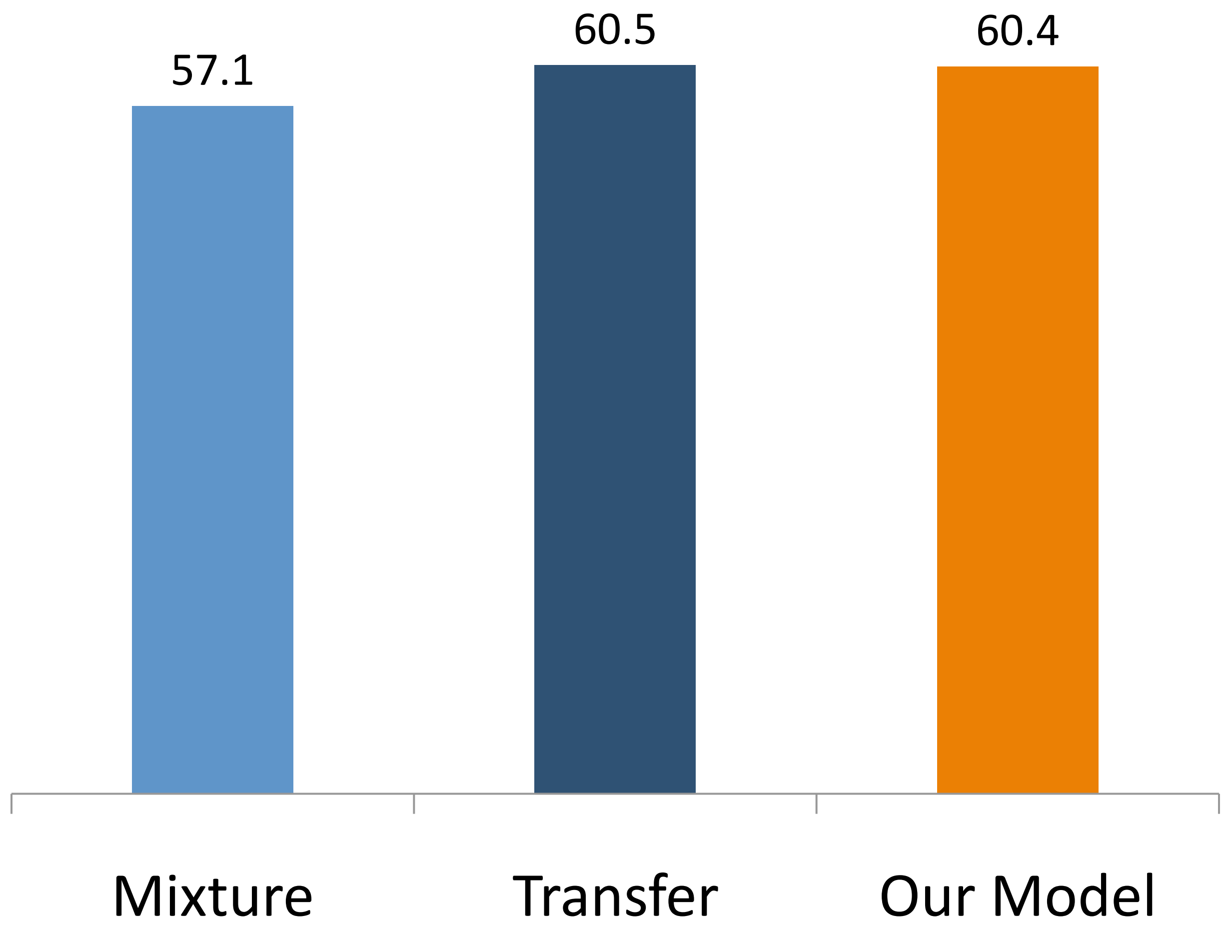
  Arabic, Basque, Bulgarian, Catalan, Chinese, Czech, Dutch, English,German, Greek, Hungarian, Italian, Japanese, Portuguese,Spanish, Swedish and Turkish

- CoNLL 2006 and 2007 datasets
- Sentences of length ≤ 50
- Baselines:
  - Transfer (McDonald et al., 2011)
  - Mixture (Cohen et al., 2011)

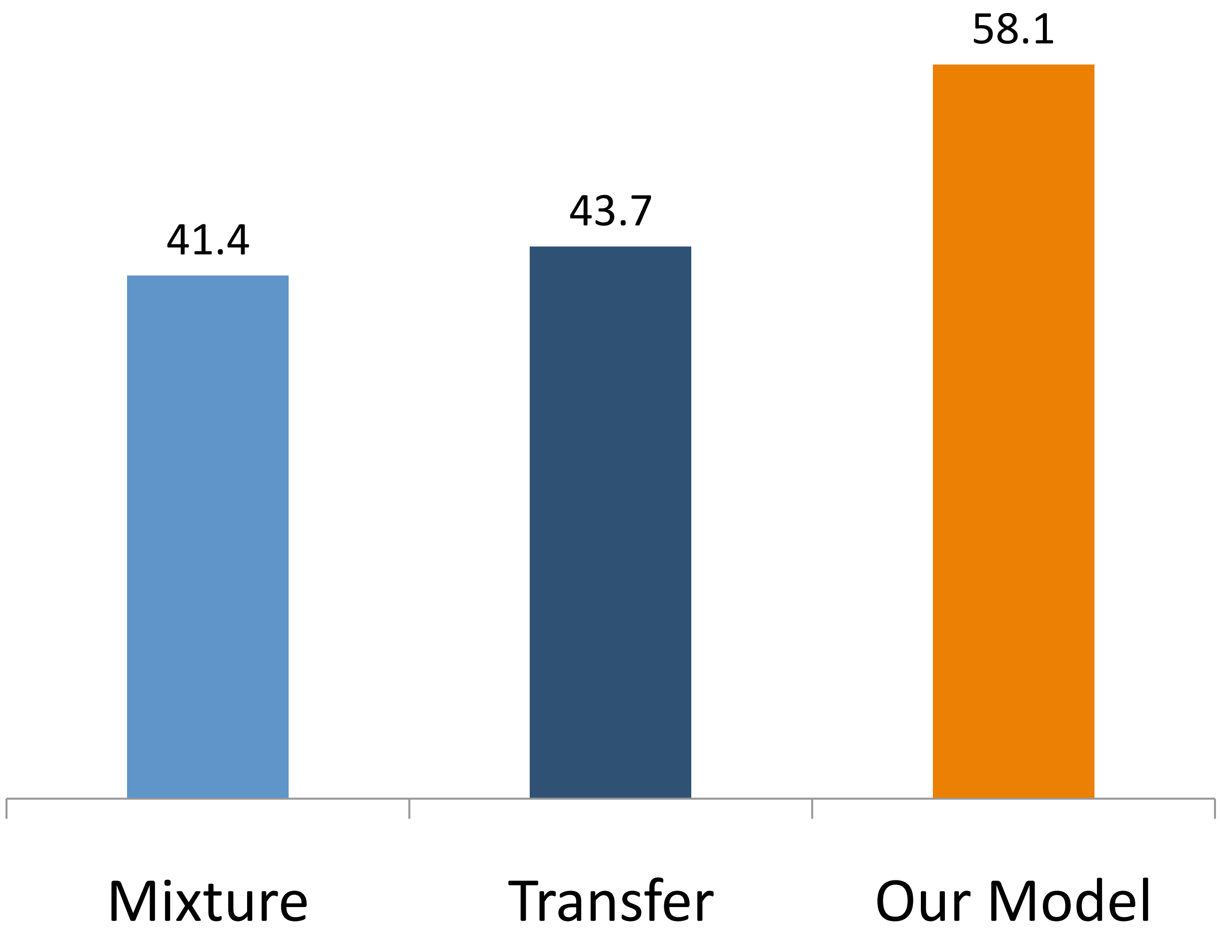
Comparison with the Baselines



Indo-European Languages



Non Indo-European Languages



Contributions

- Developed a novel generative model for multilingual parsing to incorporate linguistic knowledge
- Separated selection from ordering in the generative model in order to enable more effective parser transfer
- Designed a selective sharing mechanism for ordering parameters
- **Better dependency parsers that work across a variety of language families are crucial to robust text analysis and tree-to-tree MT systems**

Next Steps

- Lexicalization – incorporation of lexical selection parameters into the model
  – Unsupervised learning on target language
  – Transfer of lexical selection parameters from helper languages through bilingual lexicons
- Development of algorithms for learning typological features of resource-poor languages
- Testing and evaluation on focus languages