Rudolf Rosa, Zdeněk Žabokrtský
{rosa,zabokrtsky}@ufal.mff.cuni.cz

Delexicalized Cross-lingual Transfer of Statistical Syntactic Parsers for Automatic Analysis of Low-resourced Natural Languages

Charles University in Prague
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

CGI FIT ČVUT, Praha, 6 May 2016

Outline

- Introduction to linguistic analysis
- MSTParser and its delexicalization
- Single-source delexicalized parser transfer
  - KL_{cpos3} language similarity
- Multi-source delexicalized parser transfer
  - treebank concatenation
  - parse tree combination
  - model interpolation
- Future work: lexicalization

Outline

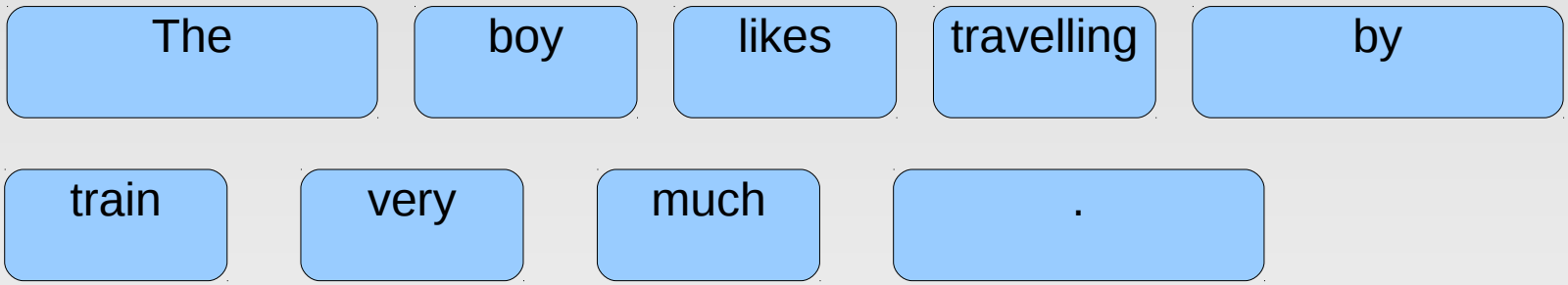
- **Introduction to linguistic analysis**
- MSTParser and its delexicalization
- Single-source delexicalized parser transfer
  - KL_{cpos3} language similarity
- Multi-source delexicalized parser transfer
  - treebank concatenation
  - parse tree combination
  - model interpolation
- Future work: lexicalization

Introduction to linguistic analysis

- The boy likes travelling by train very much.

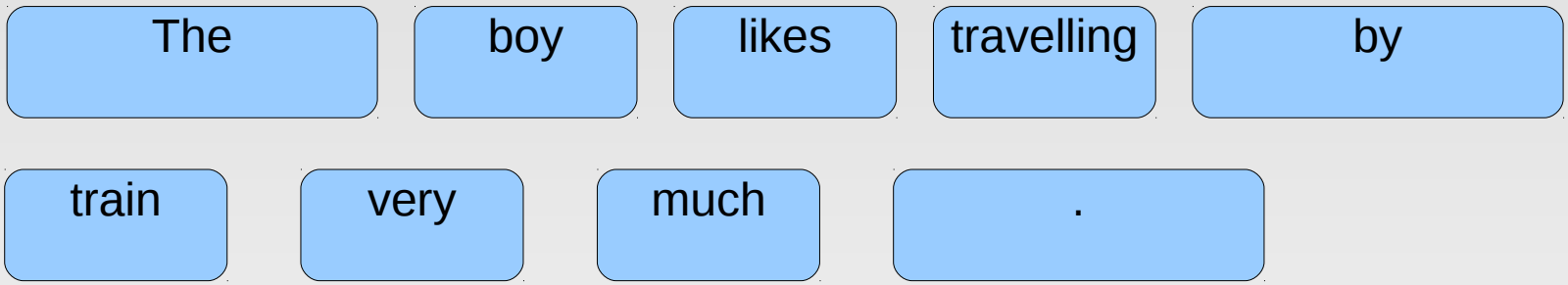
Tokenization

- The boy likes travelling by train very much.



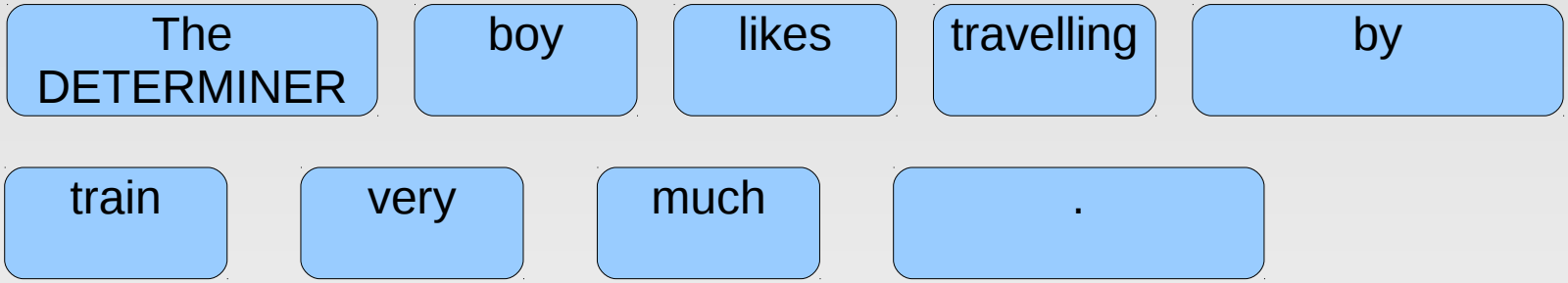
Part-of-speech tagging

- The boy likes travelling by train very much.



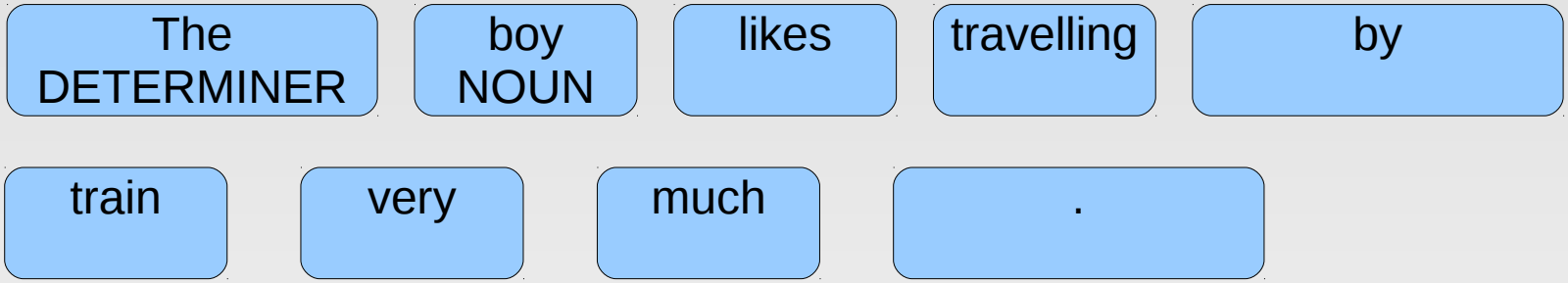
Part-of-speech tagging

- The boy likes travelling by train very much.



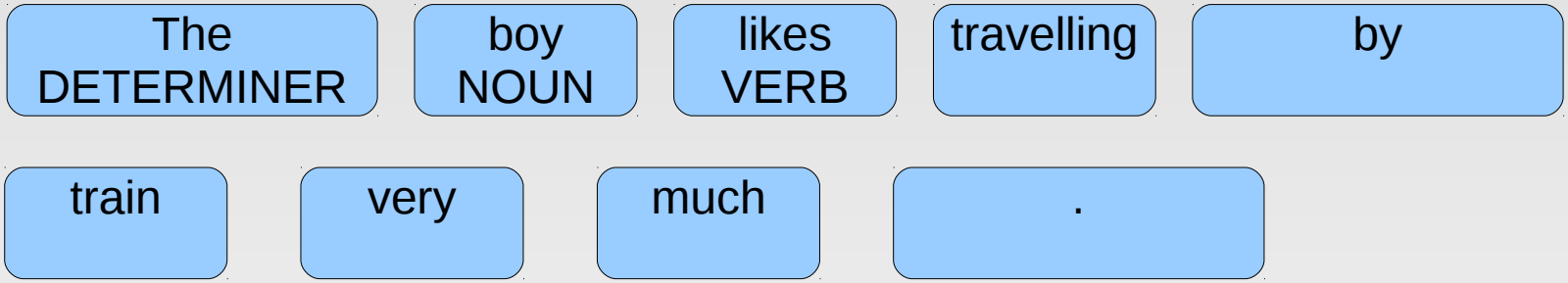
Part-of-speech tagging

- The boy likes travelling by train very much.



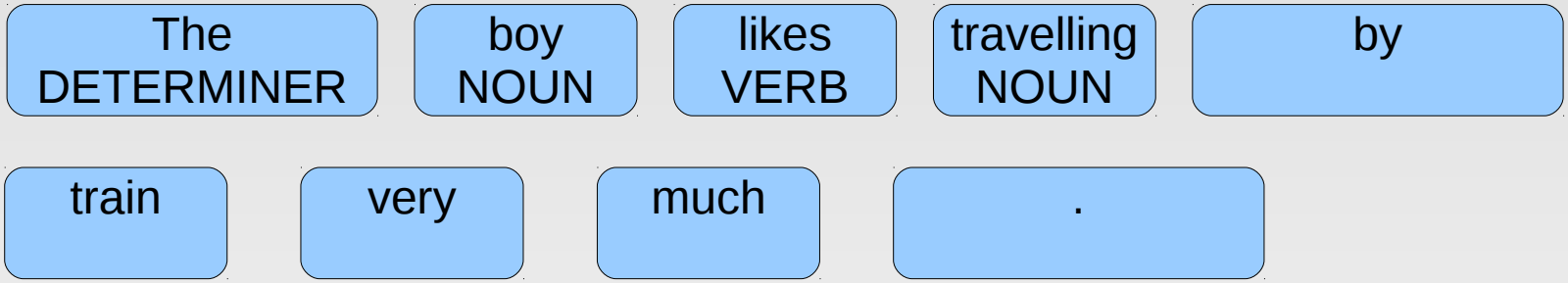
Part-of-speech tagging

- The boy likes travelling by train very much.



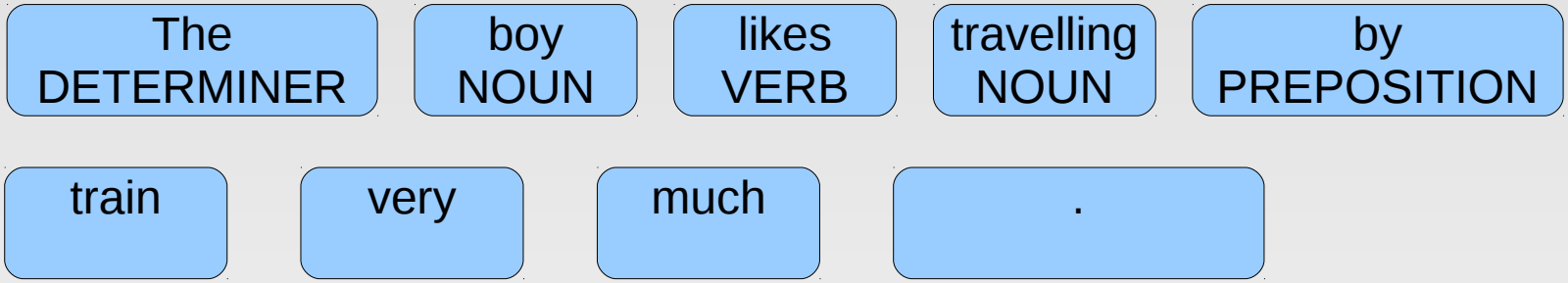
Part-of-speech tagging

- The boy likes travelling by train very much.



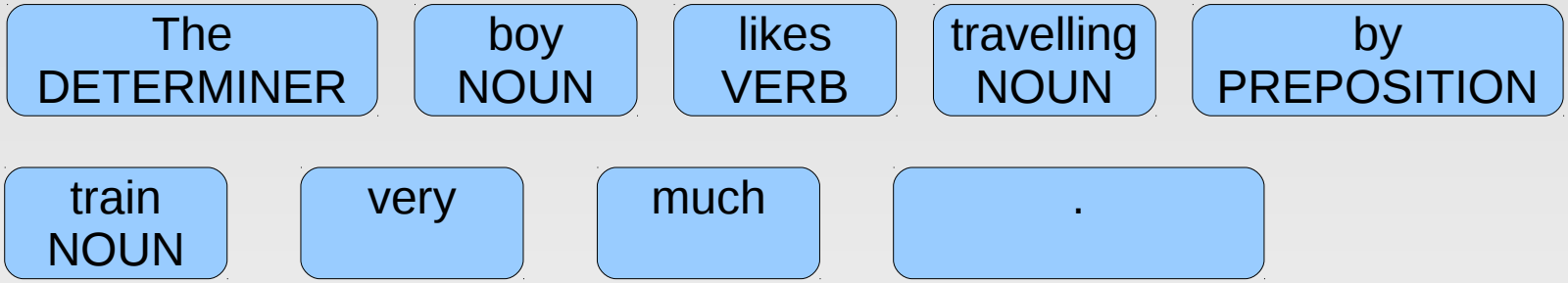
Part-of-speech tagging

- The boy likes travelling by train very much.



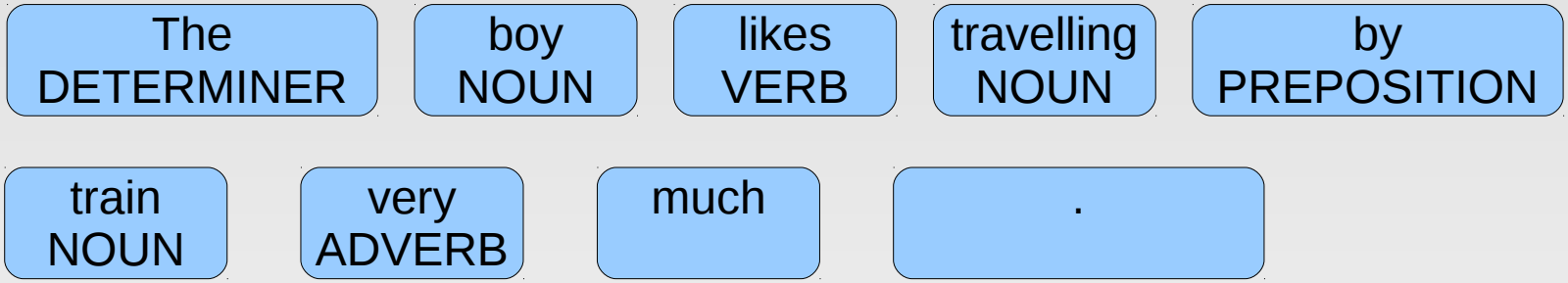
Part-of-speech tagging

- The boy likes travelling by train very much.



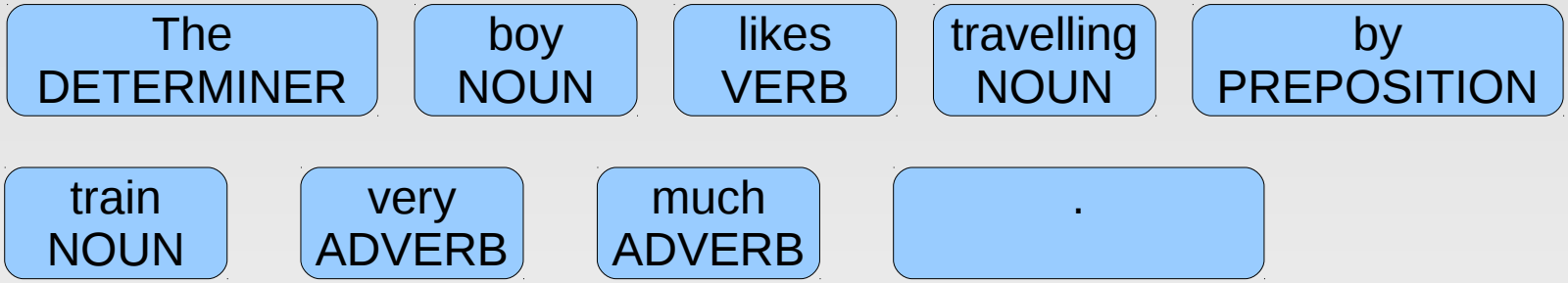
Part-of-speech tagging

- The boy likes travelling by train very much.



Part-of-speech tagging

- The boy likes travelling by train very much.



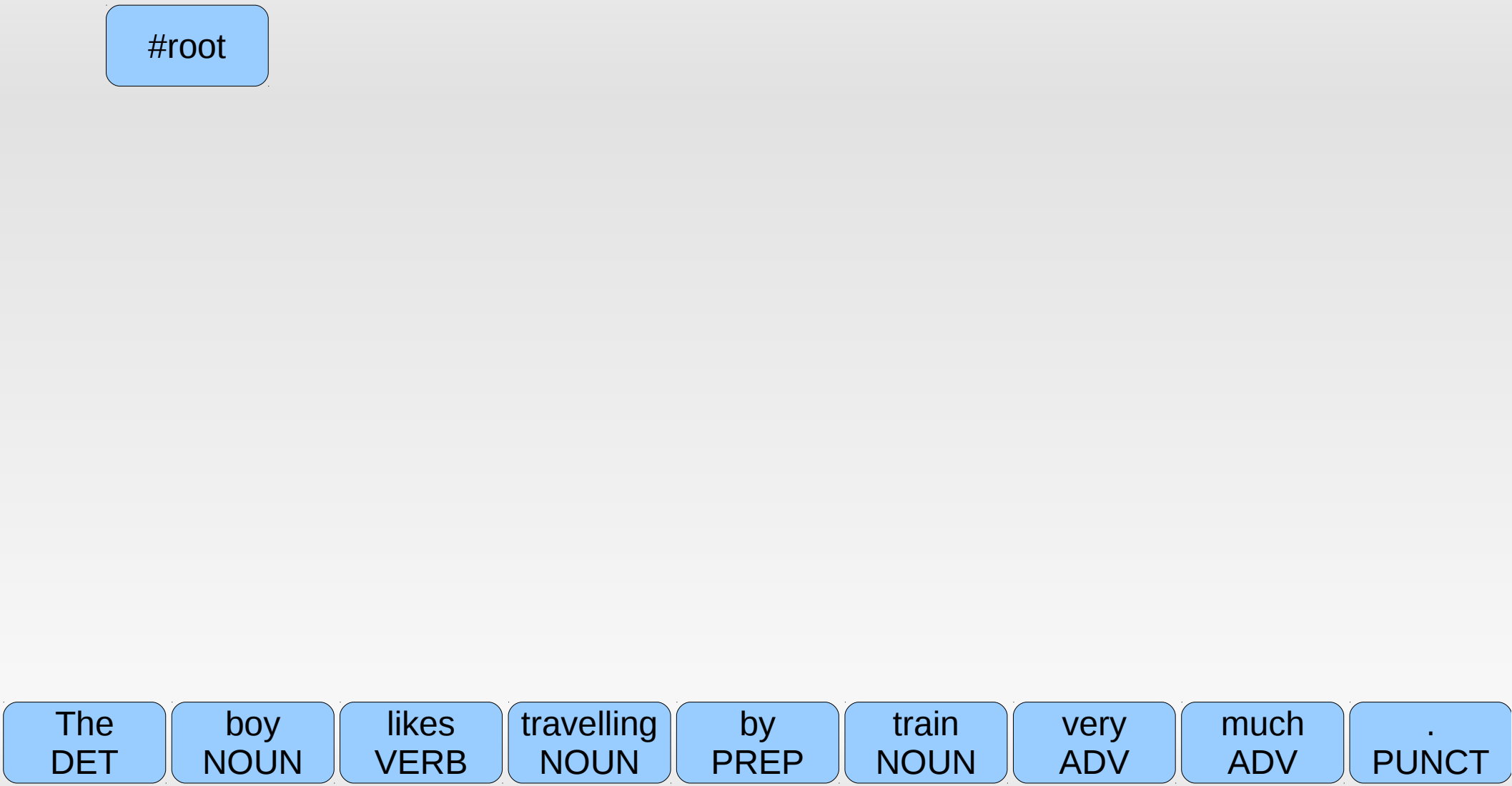
Part-of-speech tagging

- The boy likes travelling by train very much.

The DETERMINER	boy NOUN	likes VERB	travelling NOUN	by PREPOSITION

train NOUN	very ADVERB	much ADVERB	. PUNCTUATION

Syntactic parsing



Syntactic parsing

The	boy	likes	travelling	by	train	very	much
DET	NOUN	VERB	NOUN	PREP	NOUN	ADV	ADV

Syntactic parsing

The
DET

boy
NOUN

travelling
NOUN

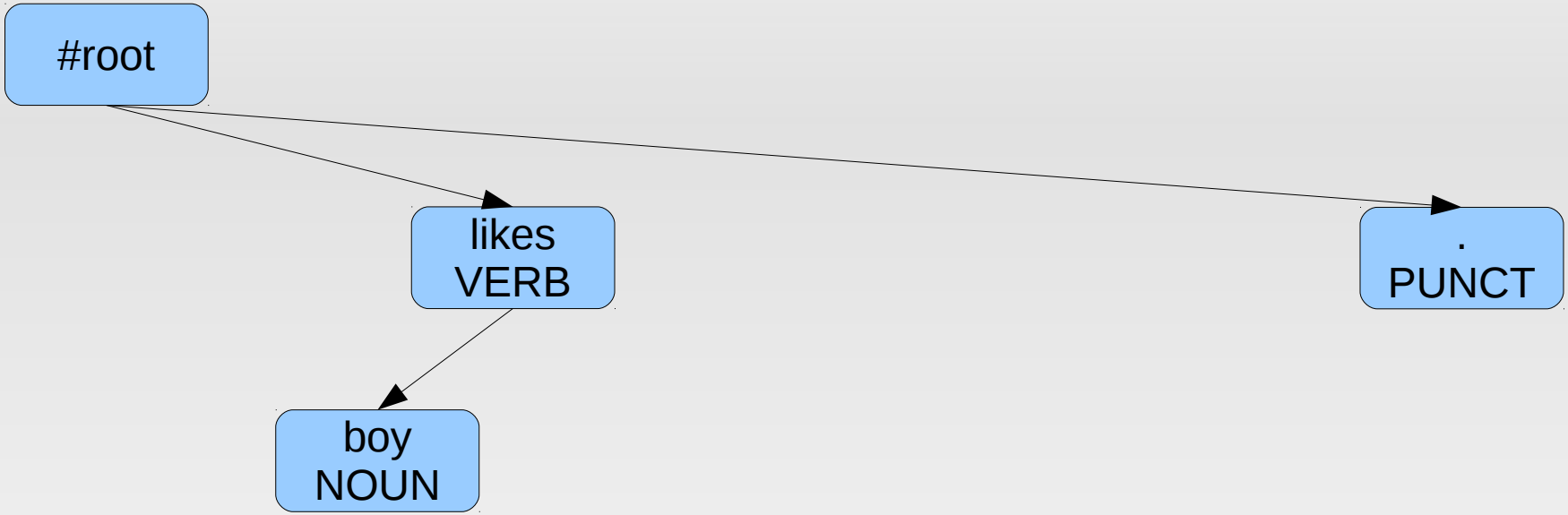
by
PREP

train
NOUN

very
ADV

much
ADV

Syntactic parsing



The
DET

travelling
NOUN

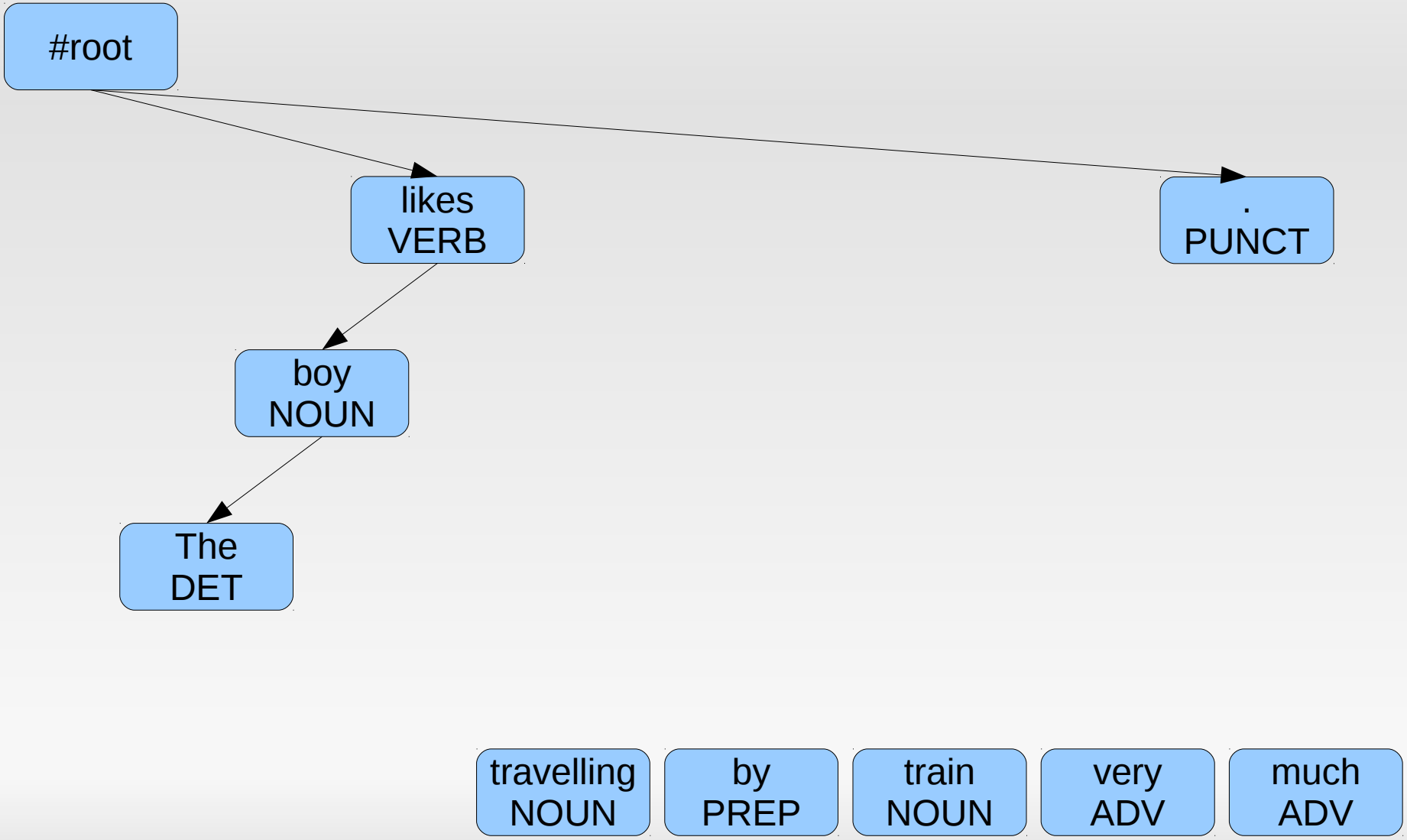
by
PREP

train
NOUN

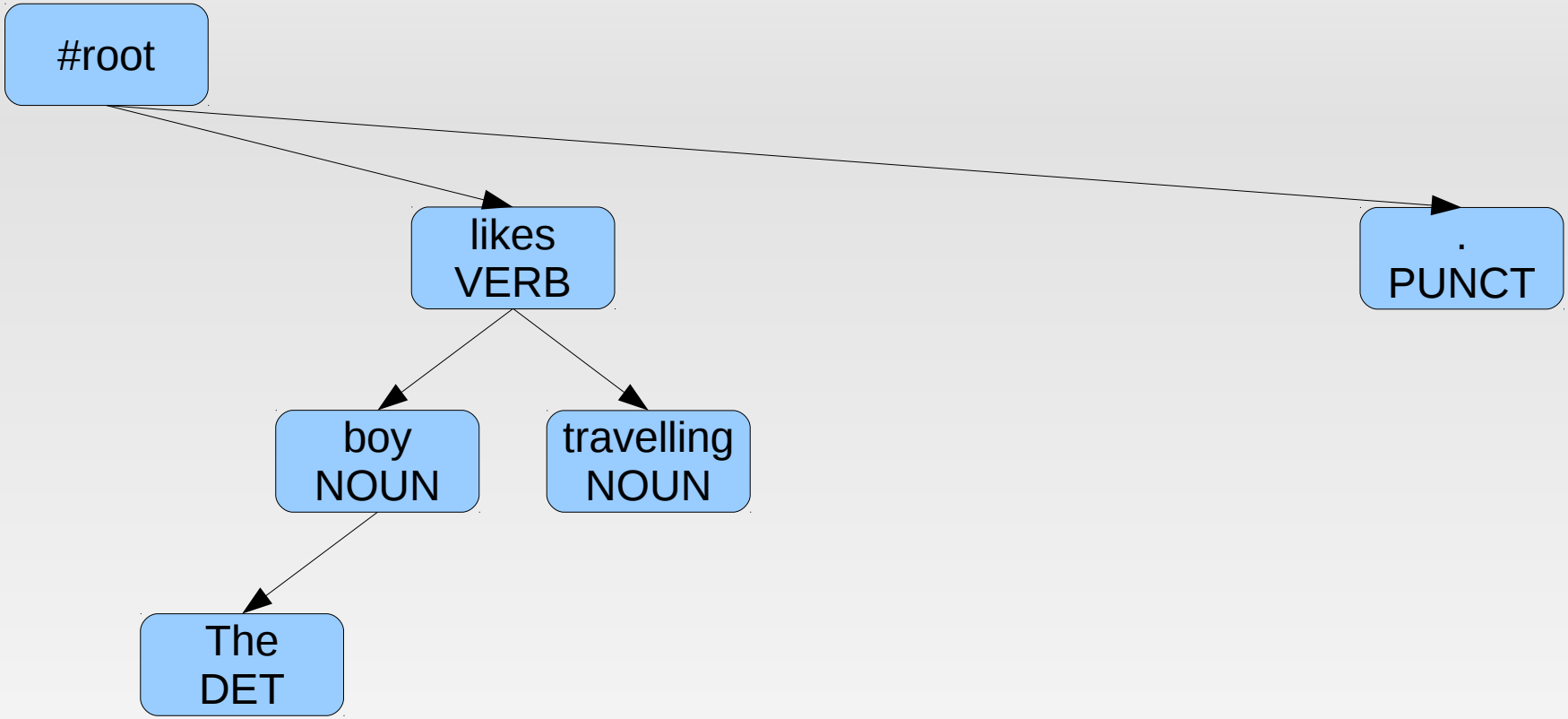
very
ADV

much
ADV

Syntactic parsing



Syntactic parsing



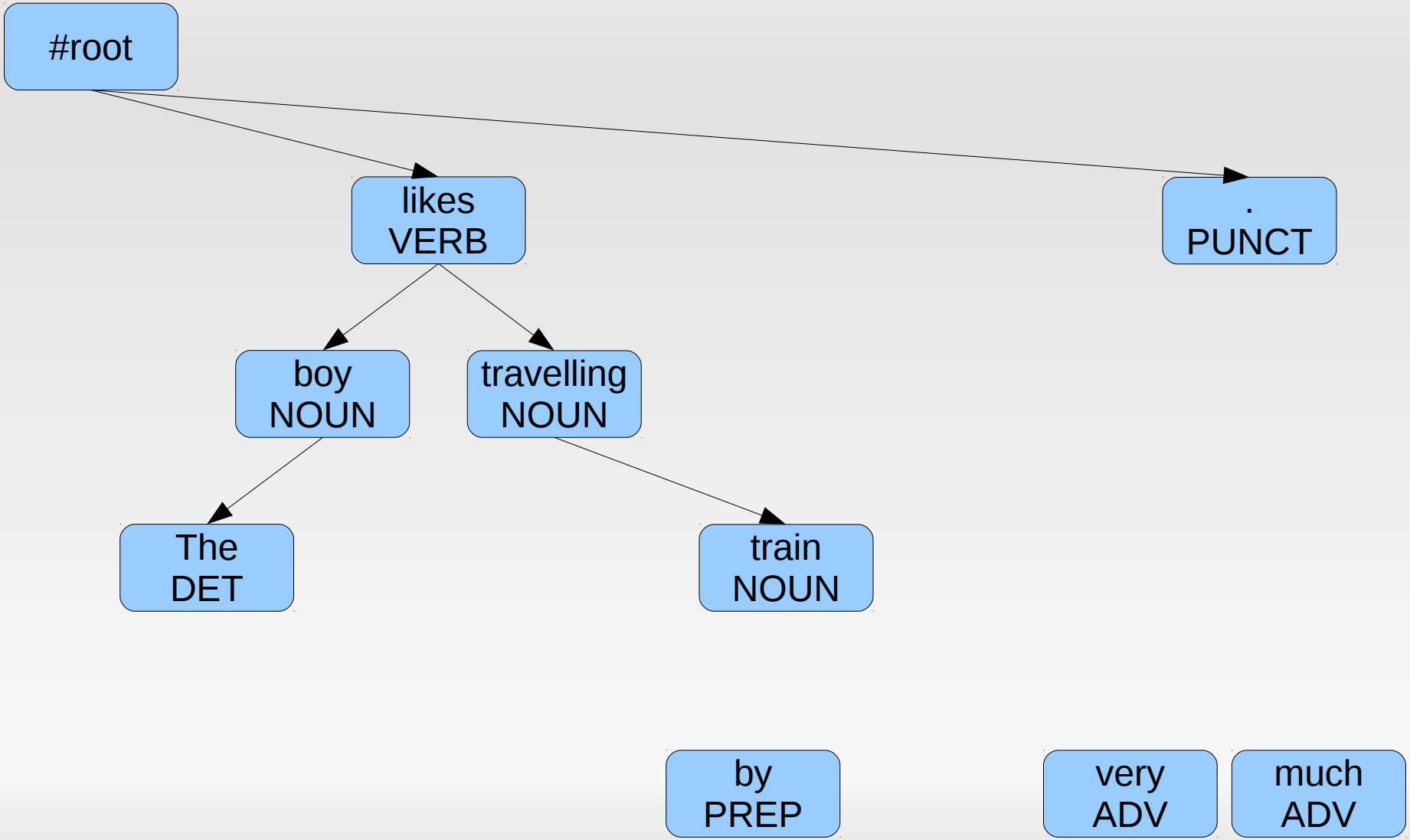
by
PREP

train
NOUN

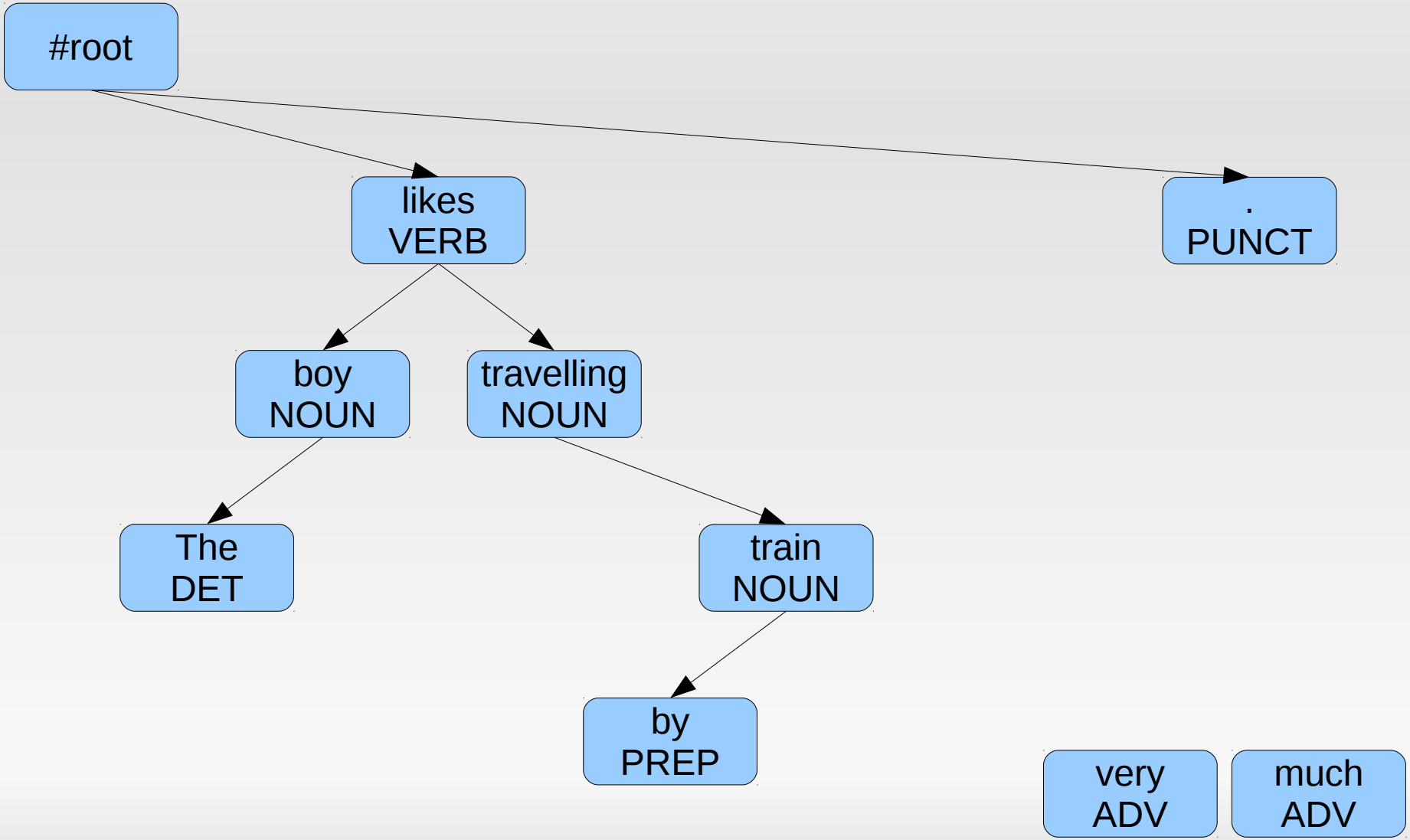
very
ADV

much
ADV

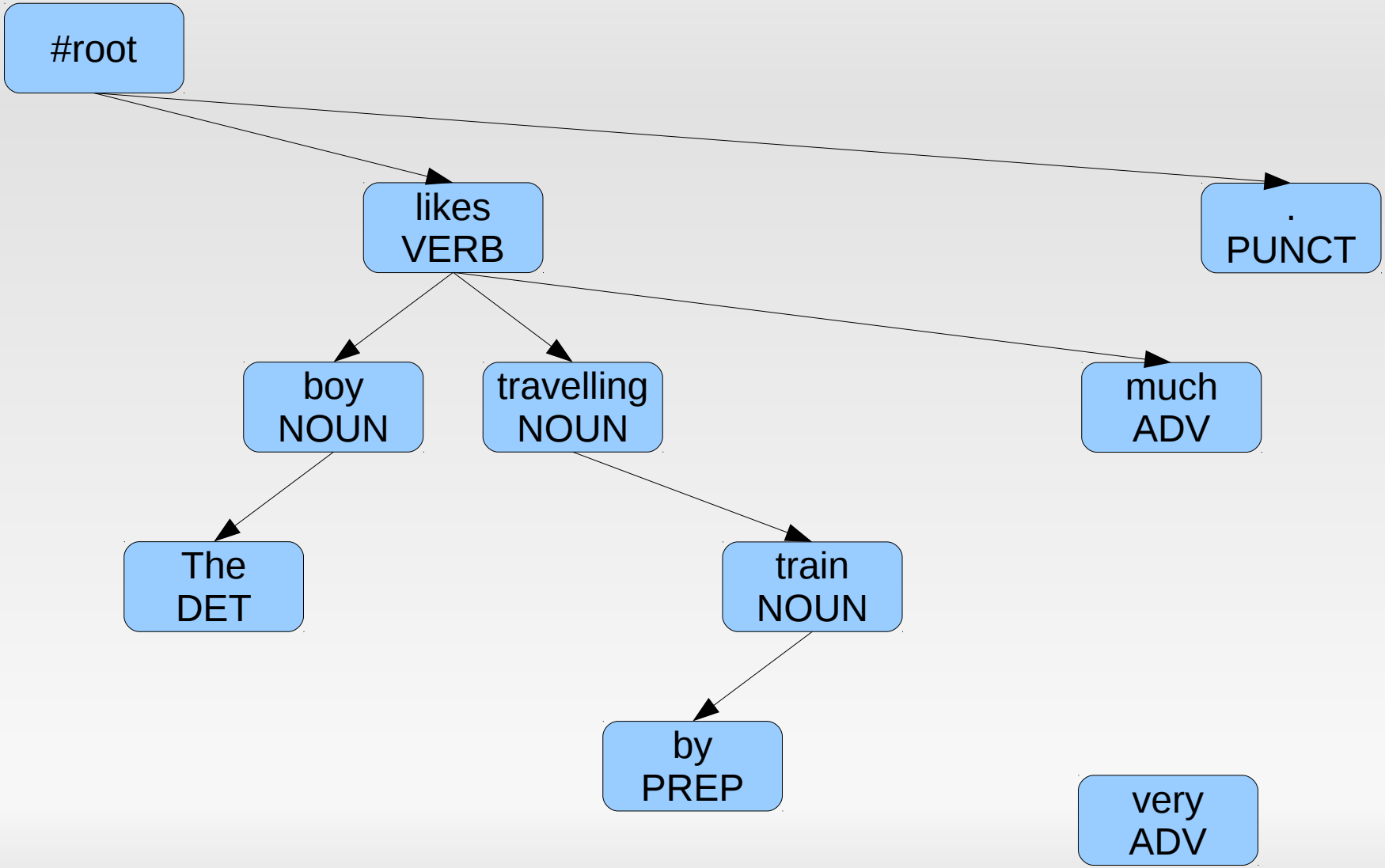
Syntactic parsing



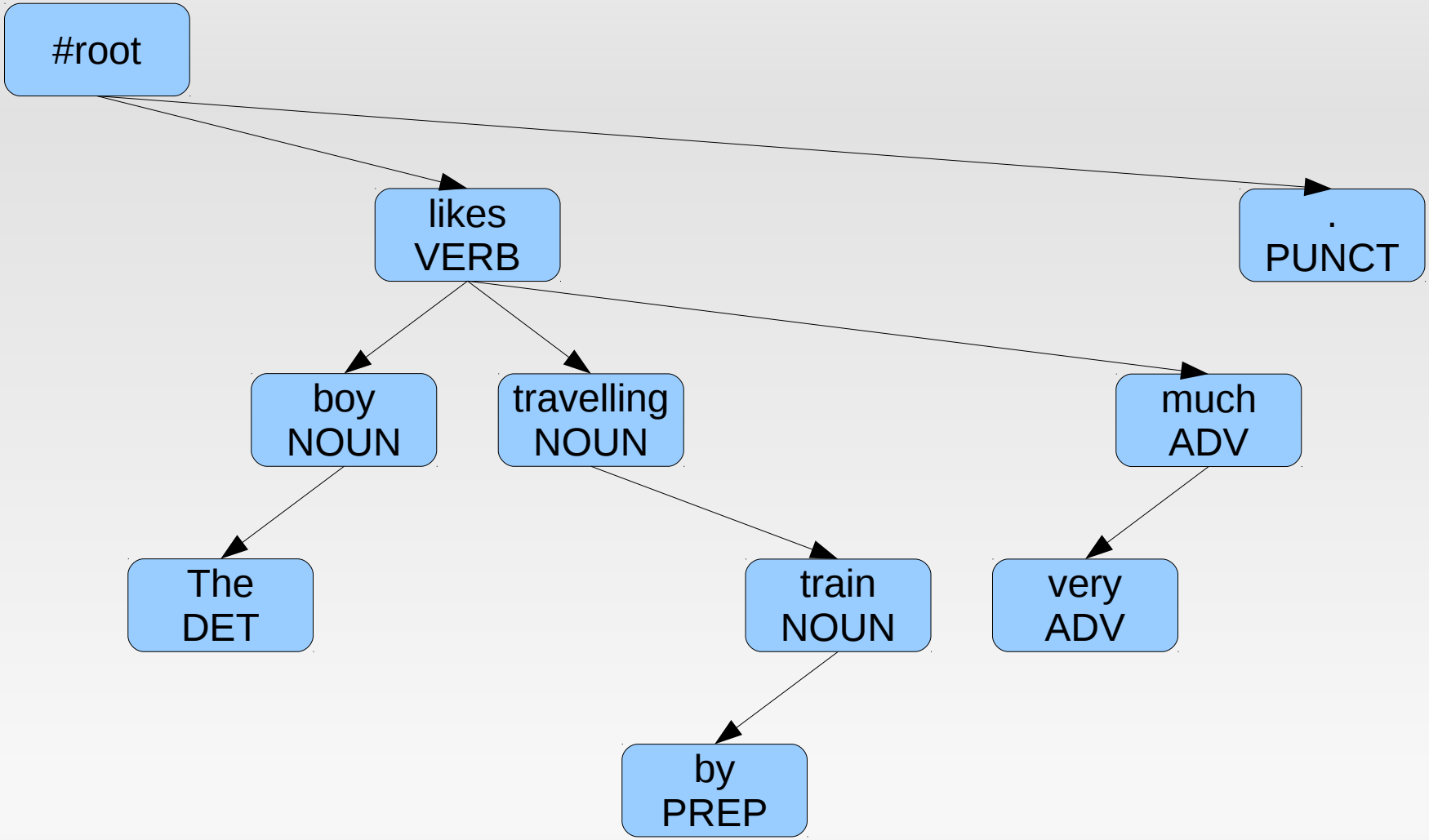
Syntactic parsing



Syntactic parsing



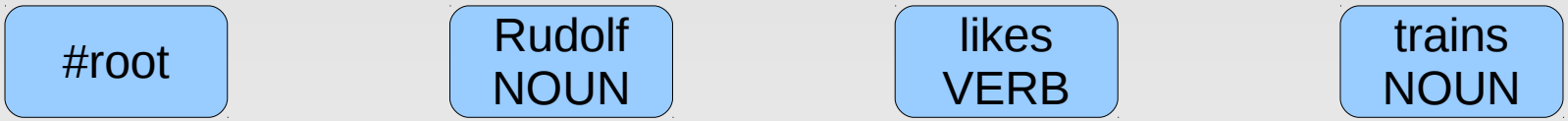
Syntactic parsing



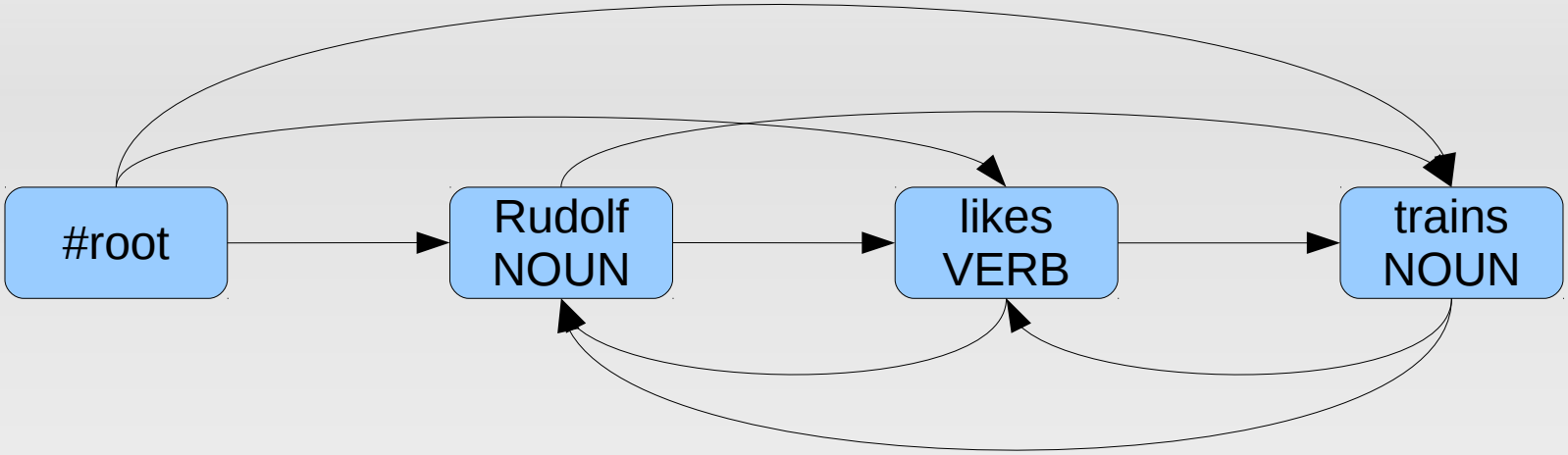
Outline

- Introduction to linguistic analysis
- **MSTParser** and its delexicalization
- Single-source delexicalized parser transfer
  - KL_{cpos3} language similarity
- Multi-source delexicalized parser transfer
  - treebank concatenation
  - parse tree combination
  - model interpolation
- Future work: lexicalization

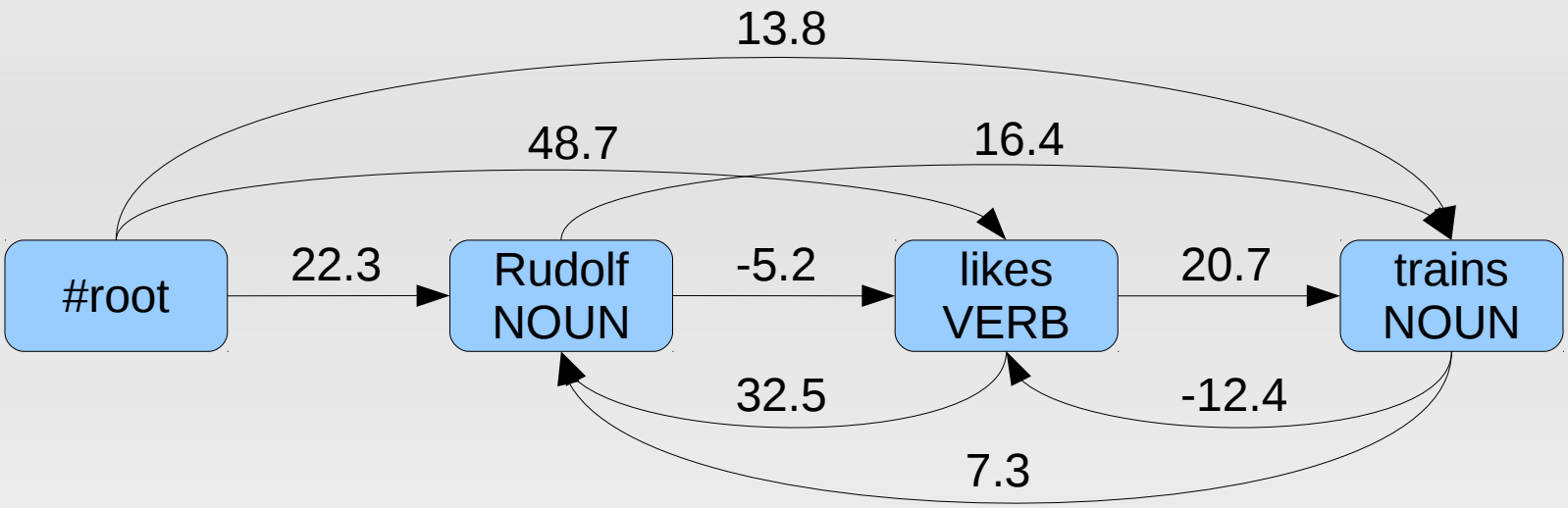
Maximum Spanning Tree Parser



Maximum Spanning Tree Parser

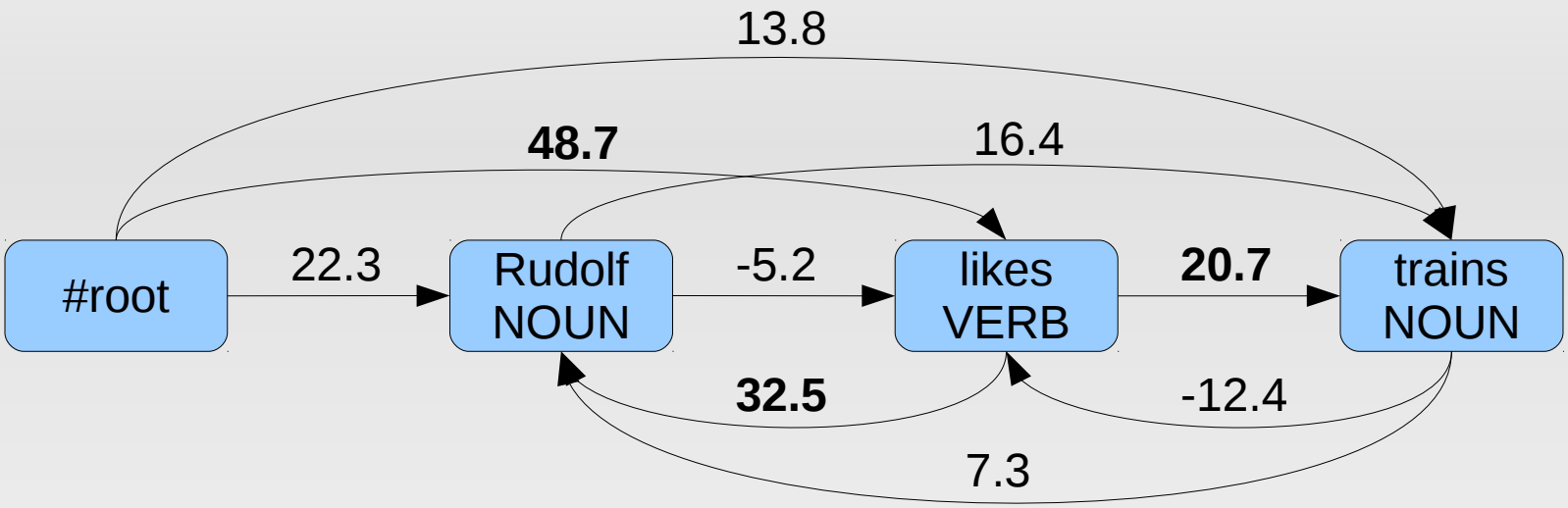


Maximum Spanning Tree Parser

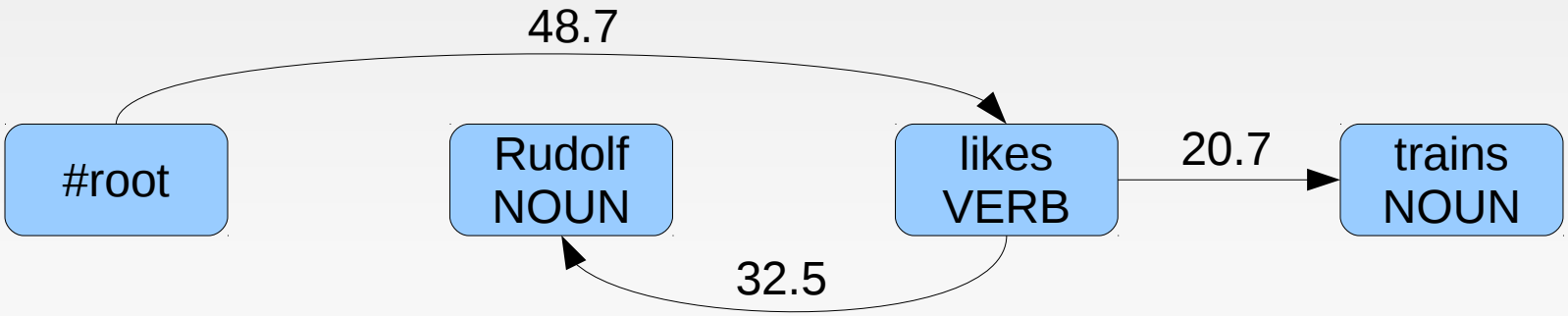


- weighting model trained on annotated data
- features – for edge nodes and their neighbours
  - lexical: word form (“likes”), word lemma (“like”)
  - morphological: part-of-speech tag (“VERB”)
  - signed distance of nodes (#root → likes: “+2”)

Maximum Spanning Tree Parser



- Chu-Liu-Edmonds MST algorithm



Outline

- Introduction to linguistic analysis
- MSTParser **and its delexicalization**
- Single-source delexicalized parser transfer
  - KL_{cpos3} language similarity
- Multi-source delexicalized parser transfer
  - treebank concatenation
  - parse tree combination
  - model interpolation
- Future work: lexicalization

Semi-supervised parsing

- fully supervised dependency parsing
  - requires training data (treebank) or a grammar

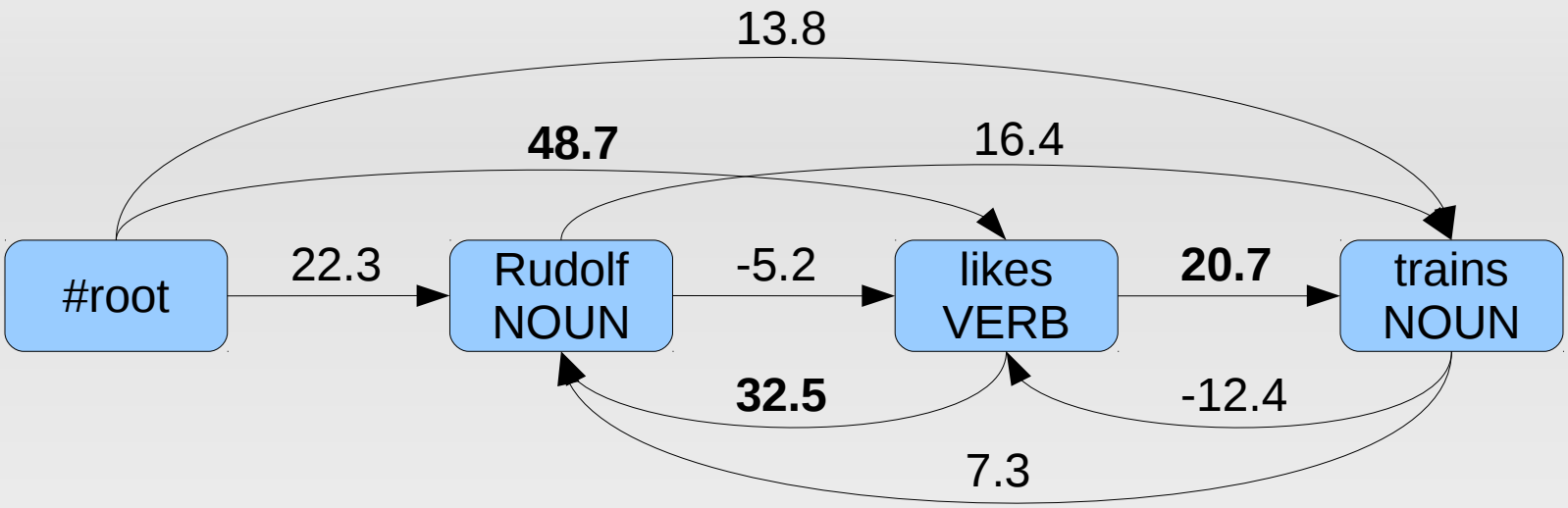
Semi-supervised parsing

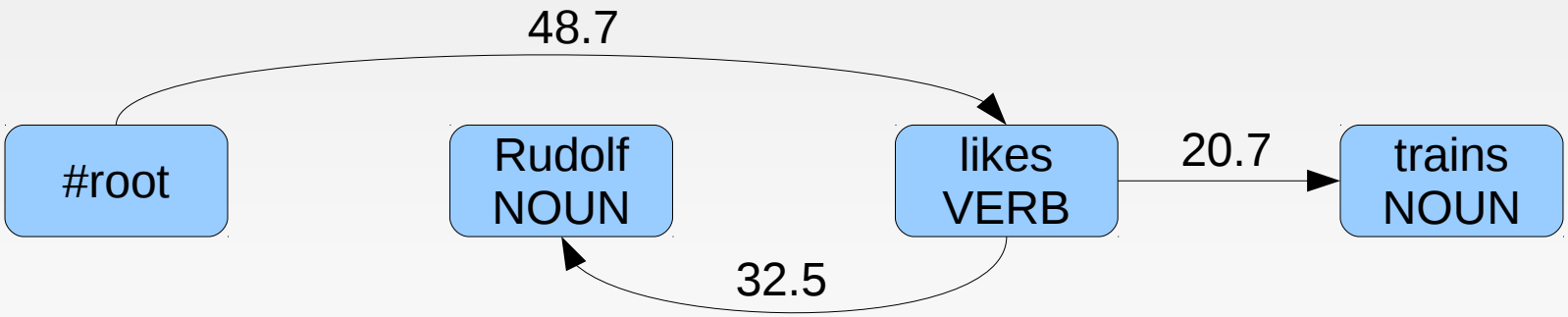
- fully supervised dependency parsing
  - requires training data (treebank) or a grammar
  - there are ~100 treebanks (manually annotated)
  - there are ~7 000 languages
  - + various domains, language evolution...

Semi-supervised parsing

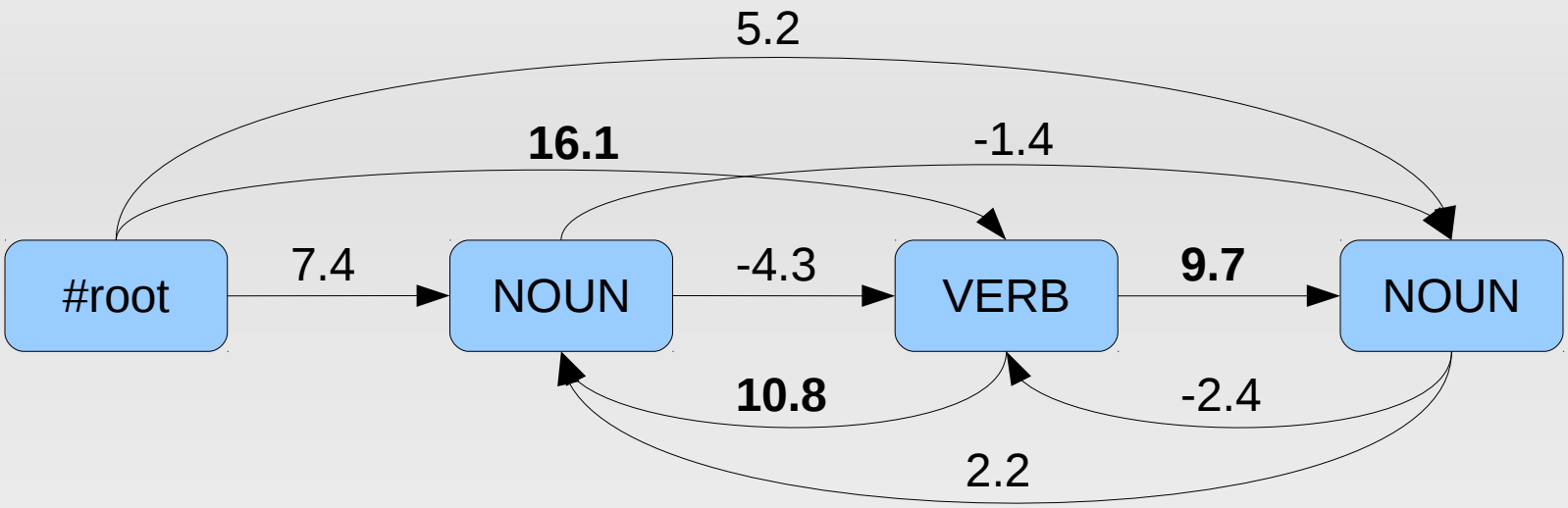
- fully supervised dependency parsing
  - requires training data (treebank) or a grammar
  - there are ~100 treebanks (manually annotated)
  - there are ~7 000 languages
  - + various domains, language evolution...
- semi-supervised parsing
  - utilize existing resources, avoid new annotations
    - treebanks for other languages (HamleDT: 30 languages)
    - unannotated data (here: part-of-speech tagged)

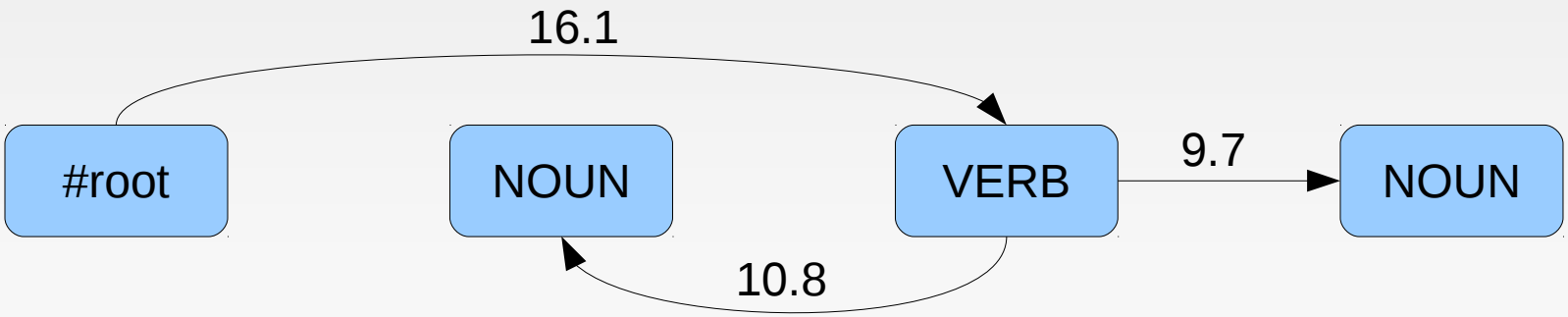
Lexicalized MSTParser





Delexicalized MSTParser





Outline

- Introduction to linguistic analysis
- MSTParser and its delexicalization
- **Single-source delexicalized parser transfer**
  - KL_{cpos3} language similarity
- Multi-source delexicalized parser transfer
  - treebank concatenation
  - parse tree combination
  - model interpolation
- Future work: lexicalization

Single-source delex parser transfer

- (Zeman and Resnik, 2008)
- train a delexicalized parser on a **source** language treebank (e.g. Czech)
- apply it to a **target** language, without a treebank but with a POS tagger (e.g. Slovak)

Utilizing multiple treebanks

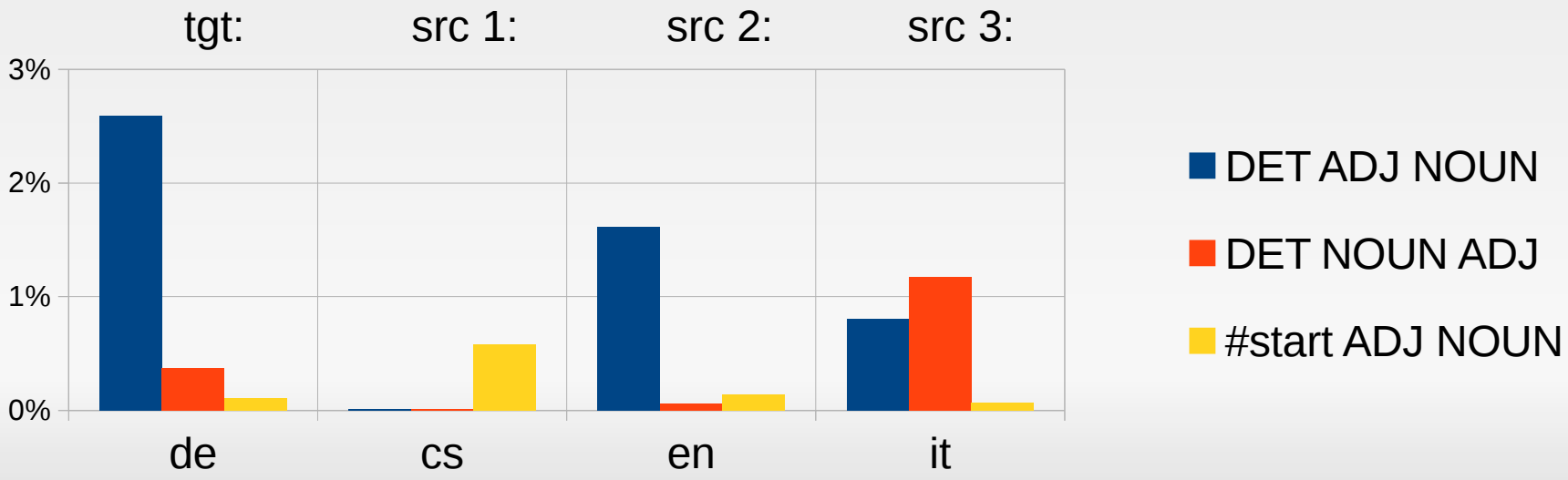
- HamleDT: 30 harmonized treebanks
- How do we choose the source treebank?
- Can we use more/all source treebanks?

Choosing the source treebank

- src should be as similar to tgt as possible
  - World Atlas of Language Structures (WALS)
    - language family, word order properties...

Choosing the source treebank

- src should be as similar to tgt as possible
  - World Atlas of Language Structures (WALS)
    - language family, word order properties...
  - $KL_{cpos\ 3}$ (***tgt***, ***src***): Kullback-Leibler divergence of POS trigram distributions



$KL_{cpos\ 3}$

$$cpos^3 = \langle cpos_{i-1}, cpos_i, cpos_{i+1} \rangle$$

$$f(cpos^3) = \frac{count(cpos^3)}{|corpus|}$$

$$KL_{cpos^3}(tgt, src) = \sum_{\forall cpos^3 \in tgt} f_{tgt}(cpos^3) \cdot \log\left(\frac{f_{tgt}(cpos^3)}{f_{src}(cpos^3)}\right)$$

Sample of results (HamleDT)

Target lang.	$KL_{cpos\ 3}$ selected src		Oracle (best possible src)
	lang.	UAS	
Bengali	Telugu	66.7	✓
Czech	Slovak	65.8	✓
Danish	Slovenian	42.1	+13.3 English
German	English	56.8	✓
Slovak	Slovenian	58.4	+ 3.3 Czech
Tamil	Turkish	31.1	+22.4 Hindi

Outline

- Introduction to linguistic analysis
- MSTParser and its delexicalization
- Single-source delexicalized parser transfer
  - KL_{cpos3} language similarity
- **Multi-source delexicalized parser transfer**
  - treebank concatenation
  - parse tree combination
  - model interpolation
- Future work: lexicalization

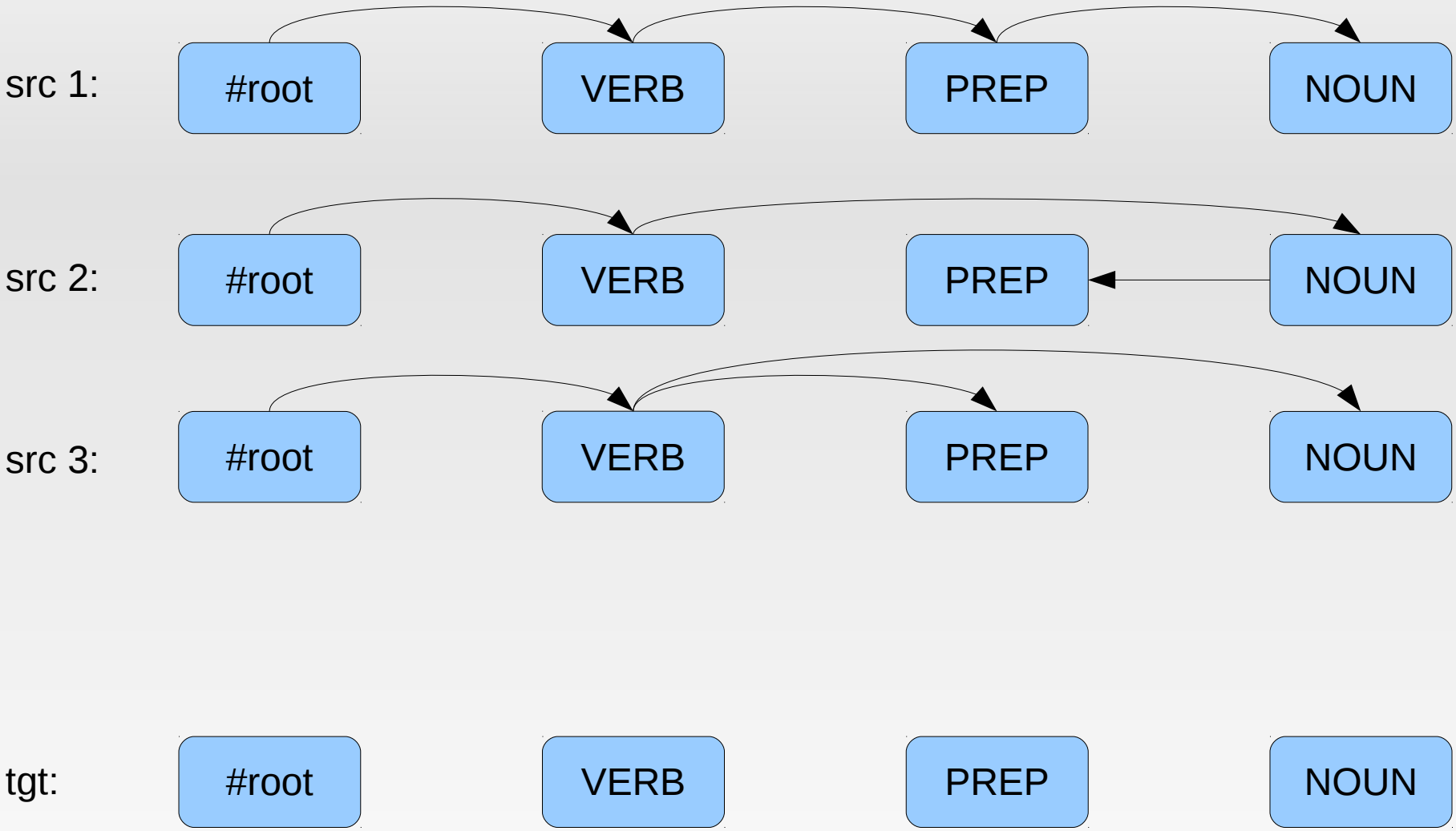
Multi-source delex parser transfer

- treebank concatenation
  - concatenate all source treebanks
  - train one delexicalized parser on the multi-treebank
  - apply the parser to the target text

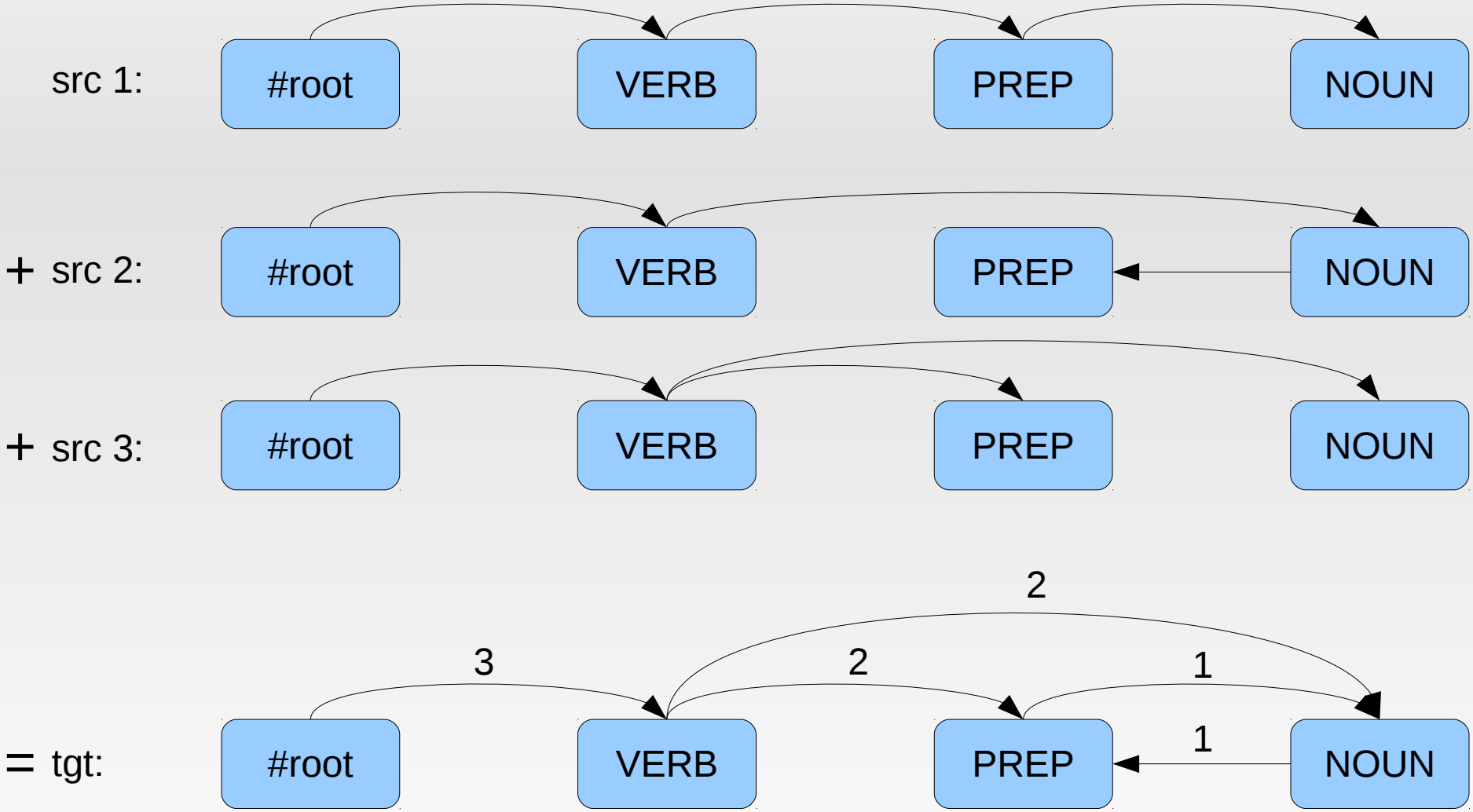
Multi-source delex parser transfer

- treebank concatenation (baseline)
  - train a parser on concatenation of all treebanks
- parse tree combination
  - train a separate parser for each source treebank
  - separately apply each parser to target text
  - use parser voting and MST algorithm to find the final analysis

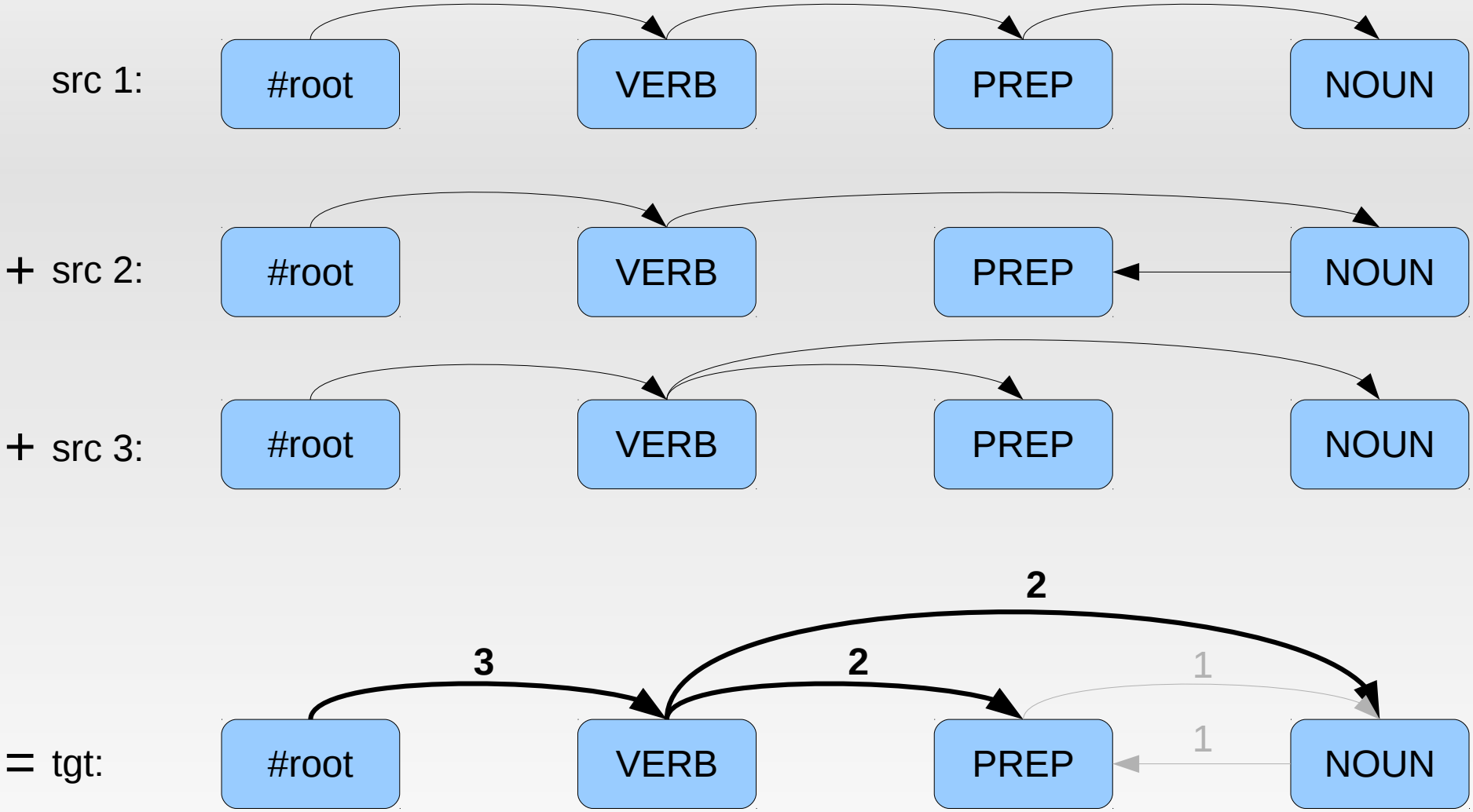
Parse tree combination



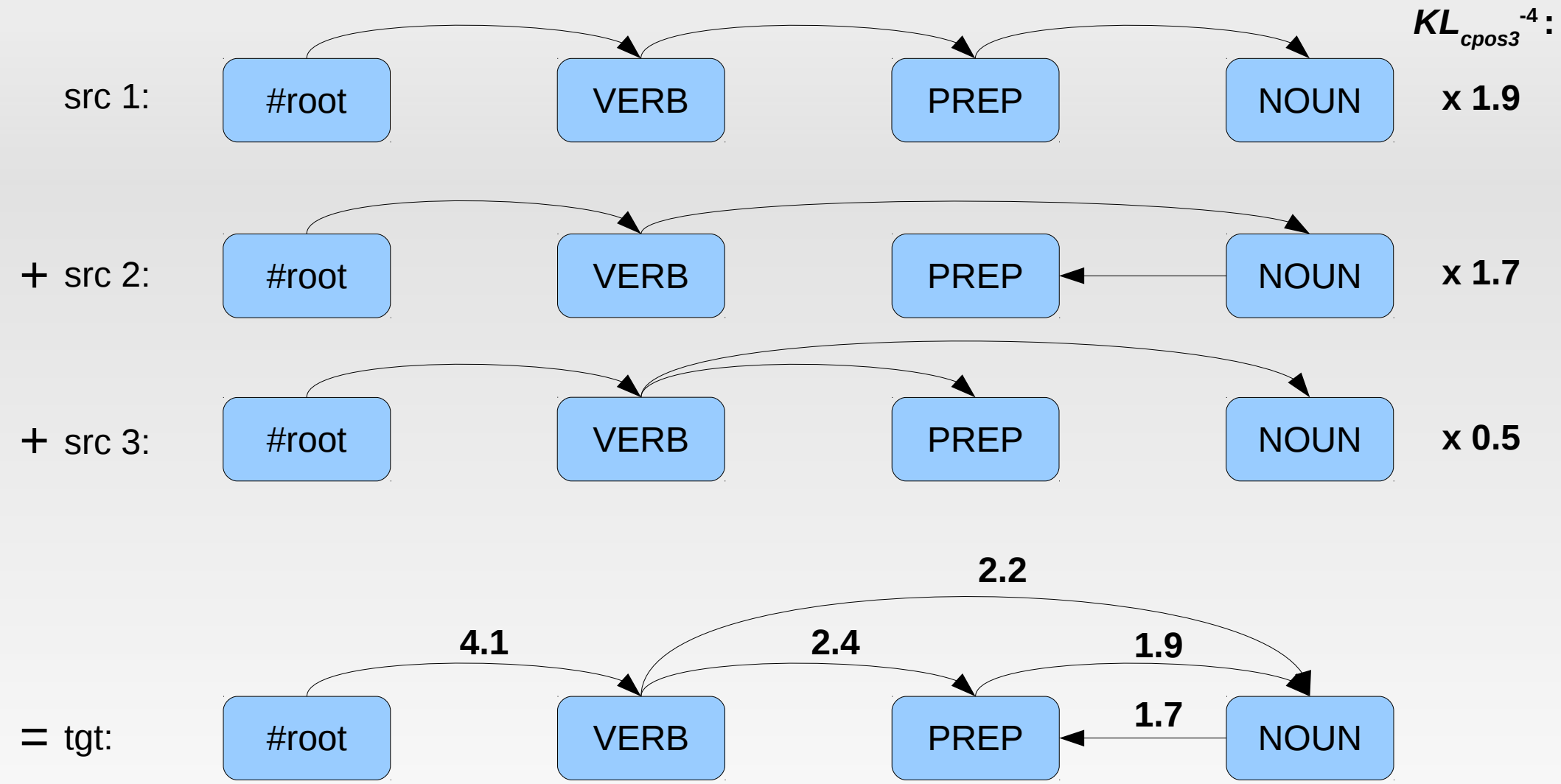
Parse tree combination



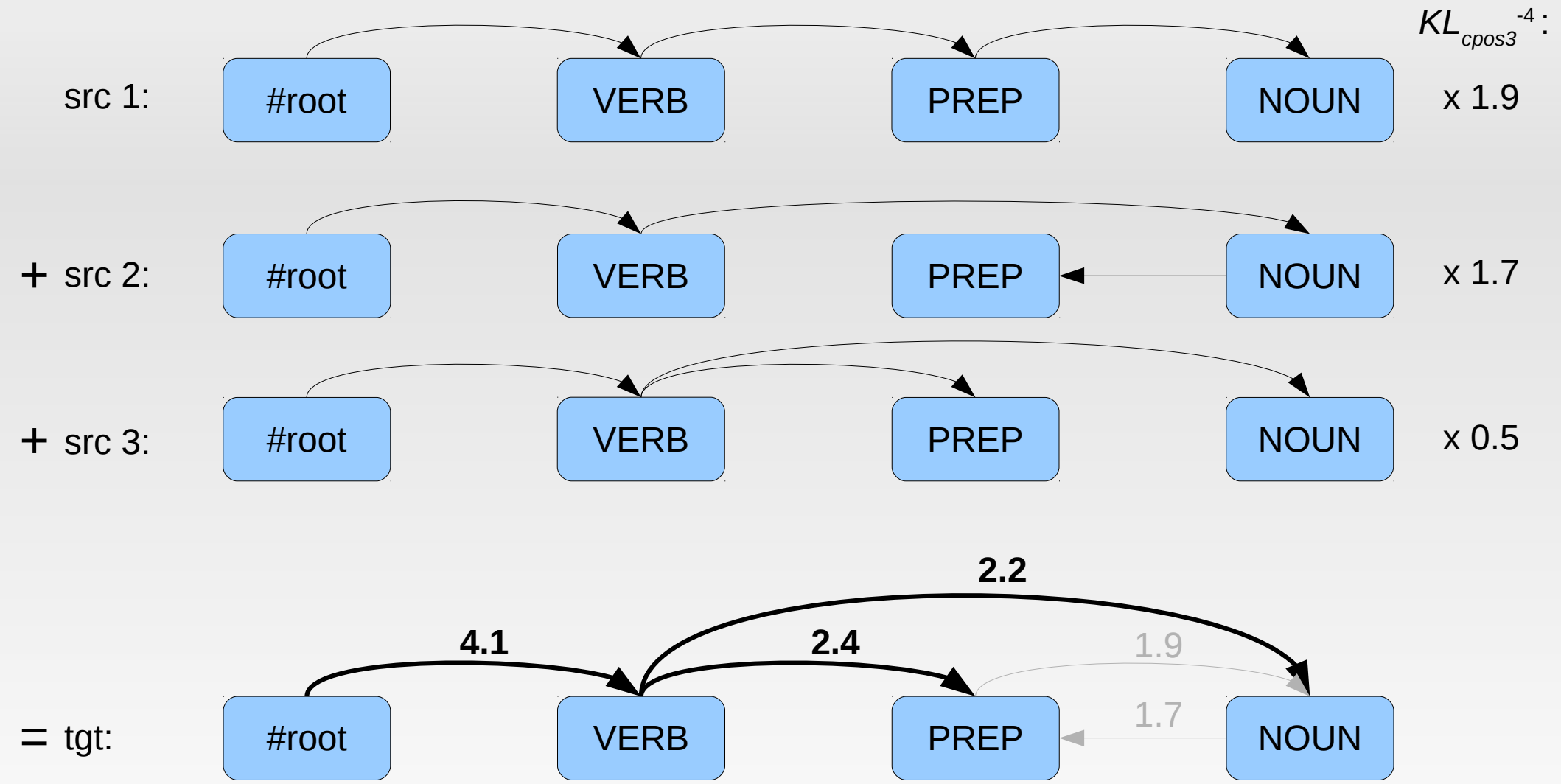
Parse tree combination



Weighted parse tree combination



Weighted parse tree combination

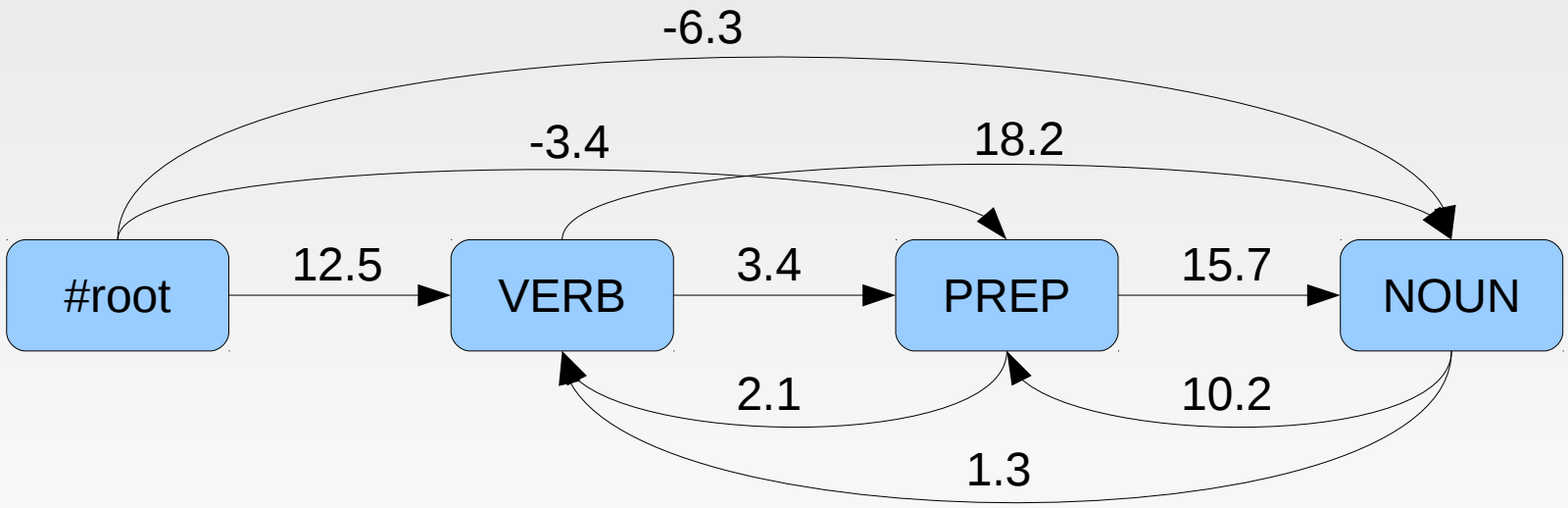


Multi-source delex parser transfer

- treebank concatenation (baseline)
  - train a parser on concatenation of all treebanks
- parse tree combination
  - combine separate src parsers via voting and MST
- parser model interpolation
  - train a parser for each source treebank
  - interpolate the trained models into a combined model
  - apply the parser with the combined model to the target text

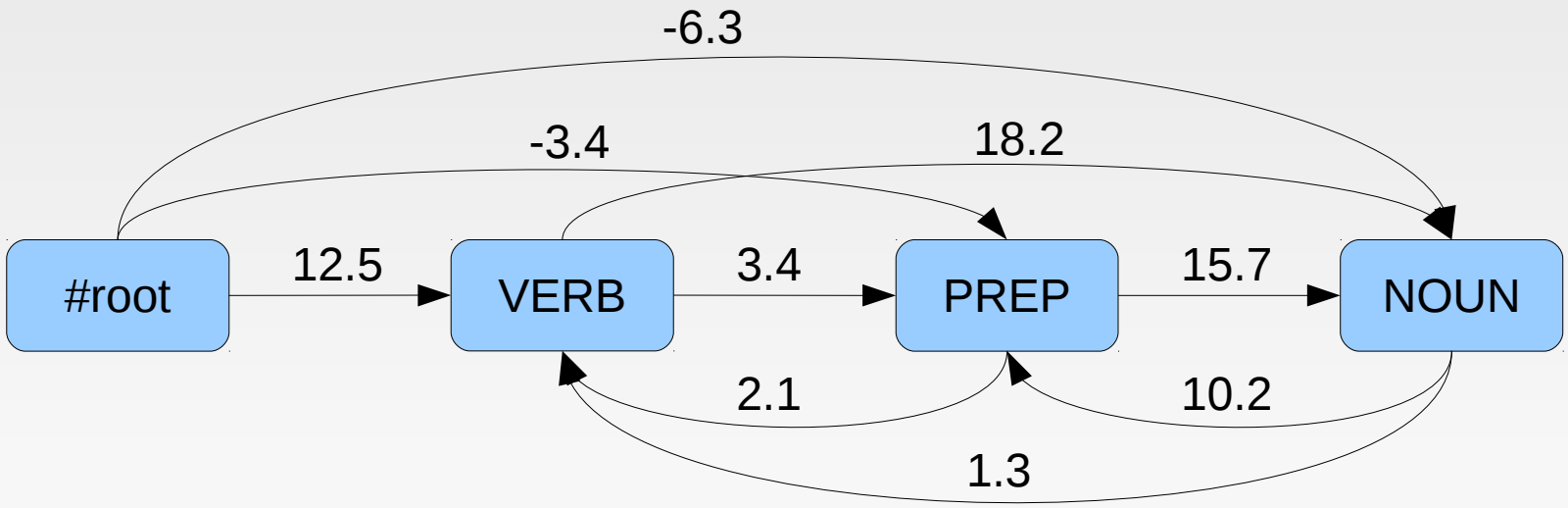
Parser model interpolation

- motivation: maybe the parser is more sure with some edges than other?
- the score assigned to the edge might show that
  - MSTParser before running the MST algorithm:

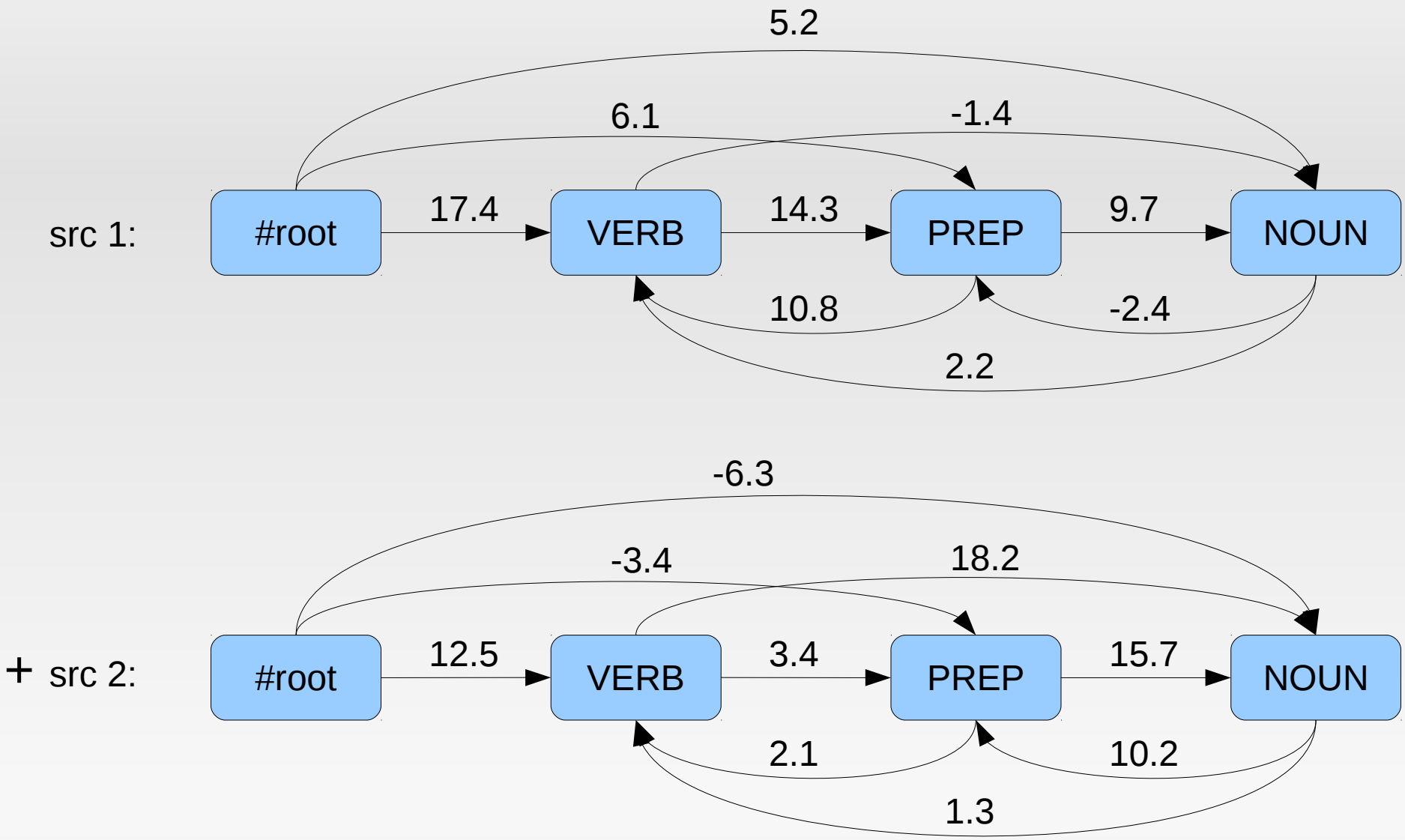


Parser model interpolation

- motivation: maybe the parser is more sure with some edges than other?
- the score assigned to the edge **might** show that
  - MSTParser before running the MST algorithm:



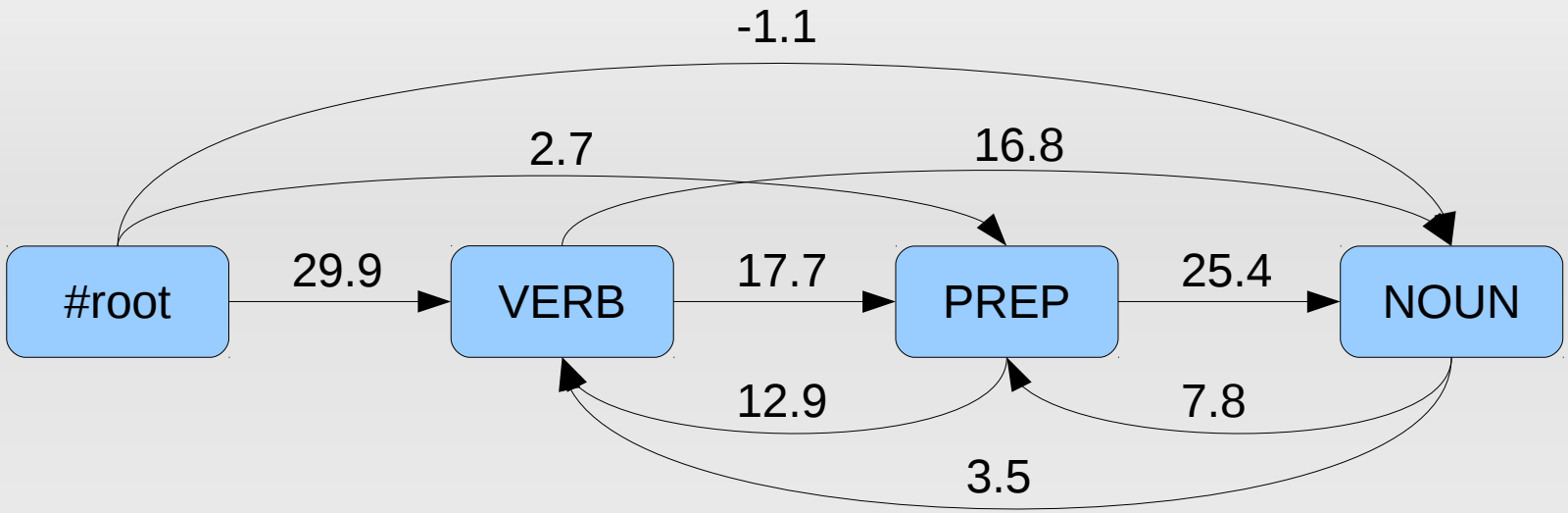
Parser model interpolation



- score normalization!

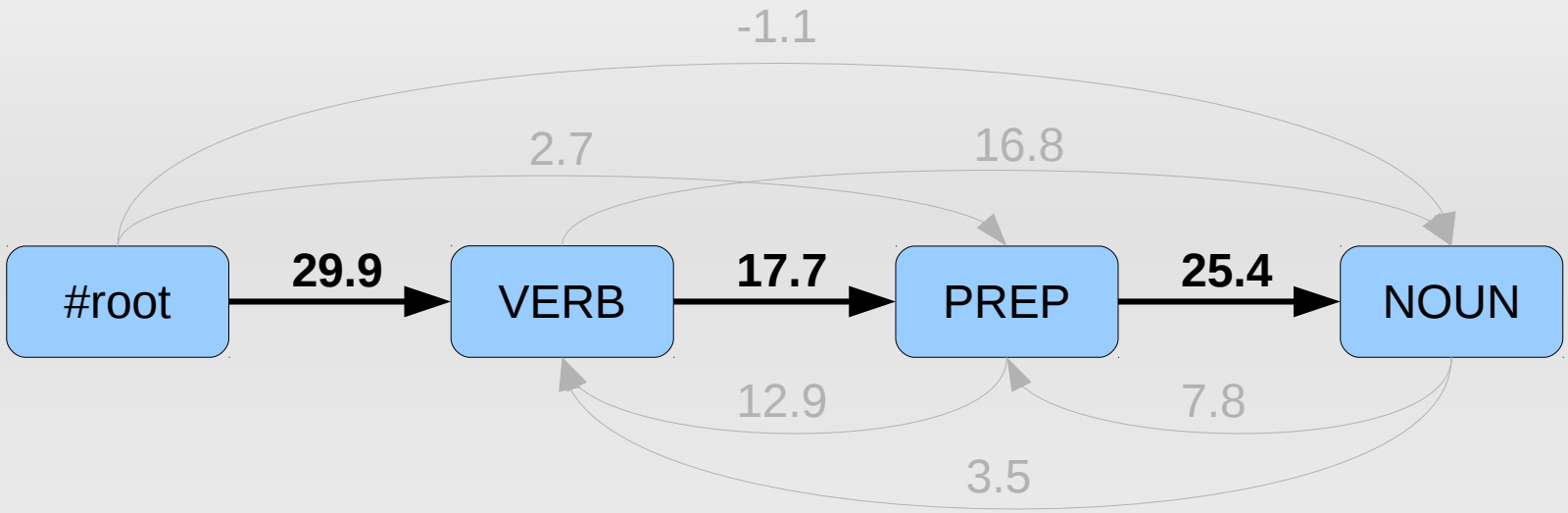
Parser model interpolation

= **tgt (Σ):**



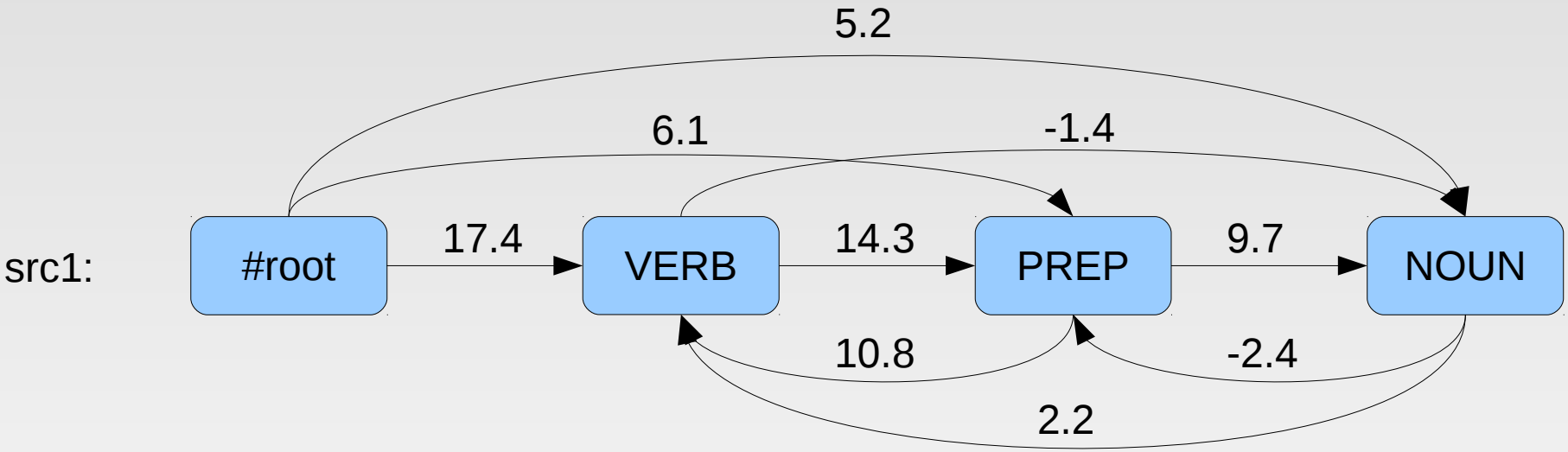
Parser model interpolation

= tgt:



Weighted parser model interpol.

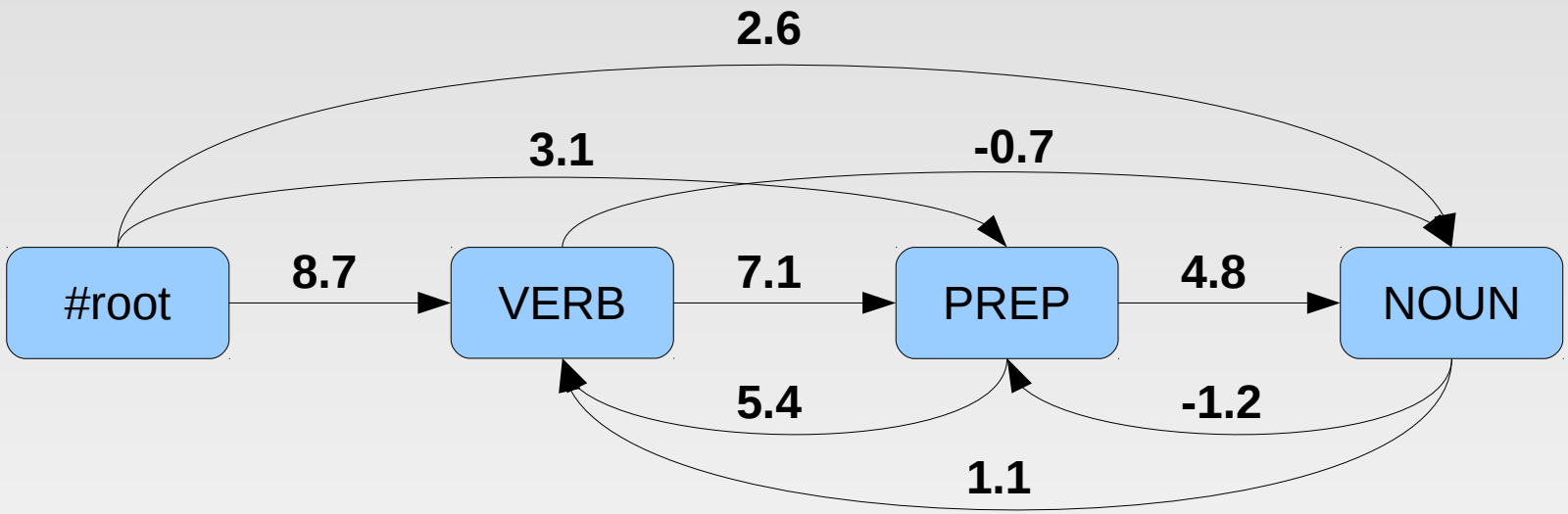
- multiply each edge score with KL_{cpos3}^{-4} (*tgt*,*src*)



KL_{cpos3}^{-4} (*tgt*, *src1*) = 0.5

Weighted parser model interpol.

- multiply each edge score with KL_{cpos3}^{-4} (*tgt*,*src*)



KL_{cpos3}^{-4} (*tgt*, *src1*) = 0.5

Why “model interpolation”?

- MSTParser edge score = **w** · **f**

Why “model interpolation”?

- MSTParser edge score = $\mathbf{w} \cdot \mathbf{f}$
- unweighted model interpolation
  - edge score = $\sum_{src} (\mathbf{w}_{src} \cdot \mathbf{f})$

Why “model interpolation”?

- MSTParser edge score = $\mathbf{w} \cdot \mathbf{f}$
- unweighted model interpolation
  - edge score = $\sum_{src} (\mathbf{w}_{src} \cdot \mathbf{f}) = (\sum_{src} \mathbf{w}_{src}) \cdot \mathbf{f}$

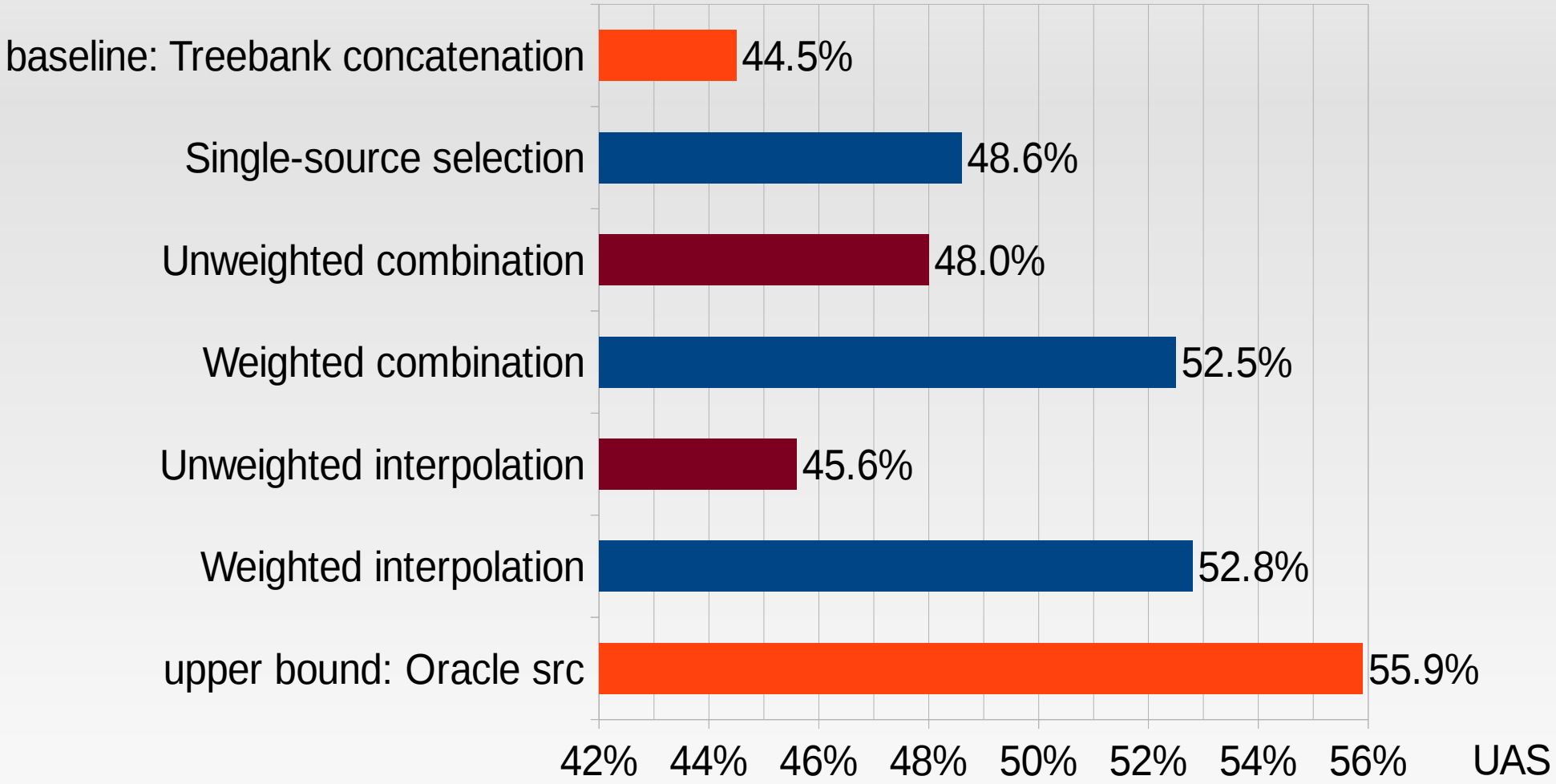
Why “model interpolation”?

- MSTParser edge score = $\mathbf{w} \cdot \mathbf{f}$
- unweighted model interpolation
  - edge score = $\sum_{src} (\mathbf{w}_{src} \cdot \mathbf{f}) = (\sum_{src} \mathbf{w}_{src}) \cdot \mathbf{f}$
  - interpolated model $\mathbf{w}_{int} = (\sum_{src} \mathbf{w}_{src})$
  - edge score = $\mathbf{w}_{int} \cdot \mathbf{f}$

Why “model interpolation”?

- MSTParser edge score = $\mathbf{w} \cdot \mathbf{f}$
- unweighted model interpolation
  - edge score = $\sum_{src} (\mathbf{w}_{src} \cdot \mathbf{f}) = (\sum_{src} \mathbf{w}_{src}) \cdot \mathbf{f}$
  - interpolated model $\mathbf{w}_{int} = (\sum_{src} \mathbf{w}_{src})$
  - edge score = $\mathbf{w}_{int} \cdot \mathbf{f}$
- weighted model interpolation: $KL_{cpos3}^{-4}(tgt, src)$
  - edge score = $\sum_{src} (KL_{src} \cdot \mathbf{w}_{src} \cdot \mathbf{f}) = (\sum_{src} KL_{src} \cdot \mathbf{w}_{src}) \cdot \mathbf{f}$
  - interpolated model $\mathbf{w}_{int} = (\sum_{src} KL_{src} \cdot \mathbf{w}_{src})$

Average UAS over 18 test TBs

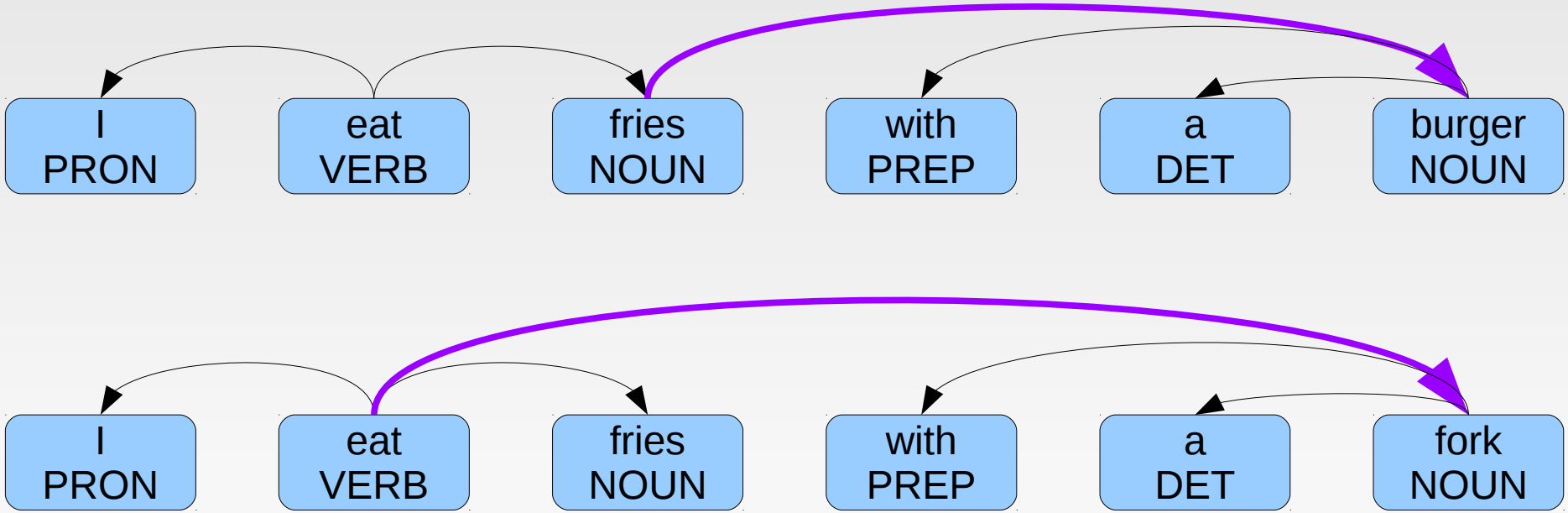


Outline

- Introduction to linguistic analysis
- MSTParser and its delexicalization
- Single-source delexicalized parser transfer
  - KL_{cpos3} language similarity
- Multi-source delexicalized parser transfer
  - treebank concatenation
  - parse tree combination
  - model interpolation
- **Future work: lexicalization**

Future work: lexicalization

- lexical features (words) are important
  - supervised parsers UAS: ~85% lex, ~75% delex
  - but they are language-specific (unlike POS tags)



Machine translation lexicalization

- parse tree transfer
  - translate target sentence to source language
  - parse translated sentence by source parser
  - transfer source parse tree to target sentence
- treebank transfer
  - translate source treebank to target language
  - transfer parse trees from source treebank to translated treebank
  - train target parser on transferred treebank

Machine translation lexicalization

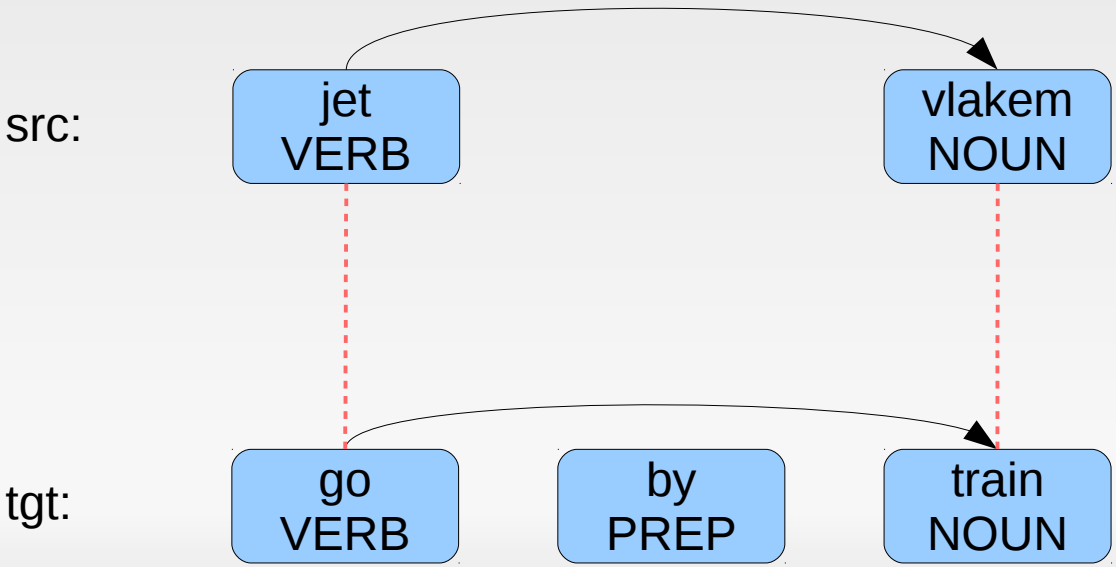
- parse tree transfer
  - transfer translated source parse tree to target sentence
- treebank transfer
  - train target parser on transferred translated source TB

Problem 1: translation

- parse tree transfer
  - transfer **translated** source parse tree to target sentence
- treebank transfer
  - train target parser on transferred **translated** source TB
- problems with machine translation
  - high-quality often available only to/from English
  - for low-resourced languages often low quality
    - requires **large** amounts of bilingual texts

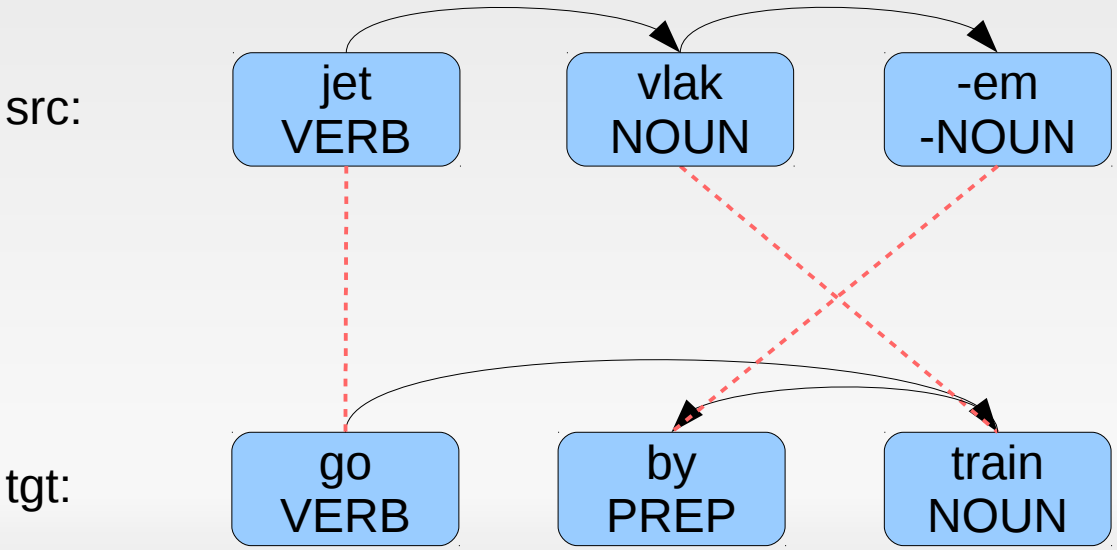
Problem 2: transfer

- parse tree transfer
  - **transfer** translated source parse tree to target sentence
- treebank transfer
  - train target parser on **transferred** translated source TB



Solution (to both): morphs?

- morphs could get closer to 1:1 correspondence
  - especially if segmentation and alignment done jointly
- translation via morphs could do with less data
  - split rare complex words into frequent simple morphs



Joint segmentation and alignment?

- given a corpus of bilingual sentences
- morph segmentation and alignment, so that
  - alignment is close to 1:1
  - aligned morphs have similar meaning
- Bayesian approach?
  - maximize probability of bilingual morphs
    - P(“-em : by”) ~ count(“-em : by”)
  - also account for alignment fertility, alignment fluency, lexical roots vs auxiliary morphs distinction

Conclusion

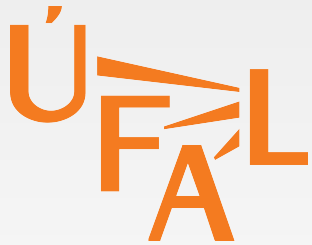
- Parsing of low-resourced natural languages
- Single-/Multi-source delexicalized parser transfer
  - parse tree combination
  - MSTParser model interpolation
  - KL_{cpos3}: language similarity for src selection/weighting
- Future work: Lexicalization
  - machine translation
  - morph splitting and alignment

Thank you for your attention

Rudolf Rosa, Zdeněk Žabokrtský
{rosa,zabokrtsky}@ufal.mff.cuni.cz

Delexicalized Cross-lingual Transfer of Statistical Syntactic Parsers for Automatic Analysis of Low-resourced Natural Languages

Charles University in Prague
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics



http://ufal.mff.cuni.cz/rudolf-rosa/