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

MSTParser Model Interpolation for Multi-source Delexicalized Transfer

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

IWPT, Bilbao, 22 July 2015

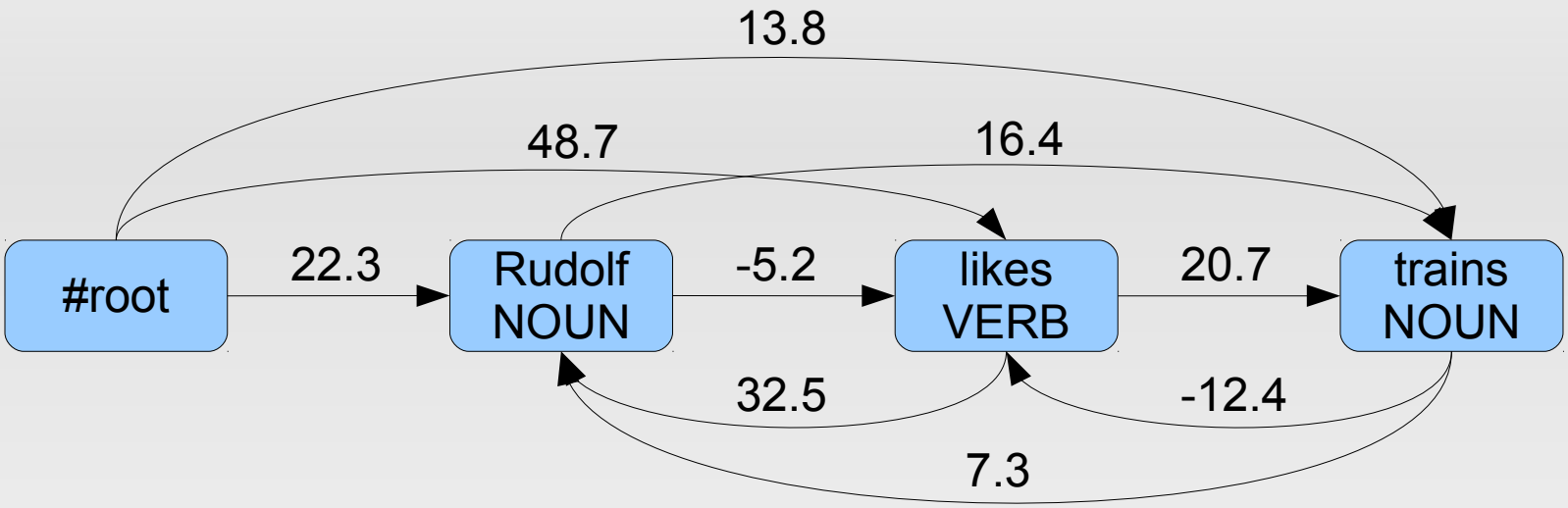
Outline

- Intro and motivation
- 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**
- Results and discussion

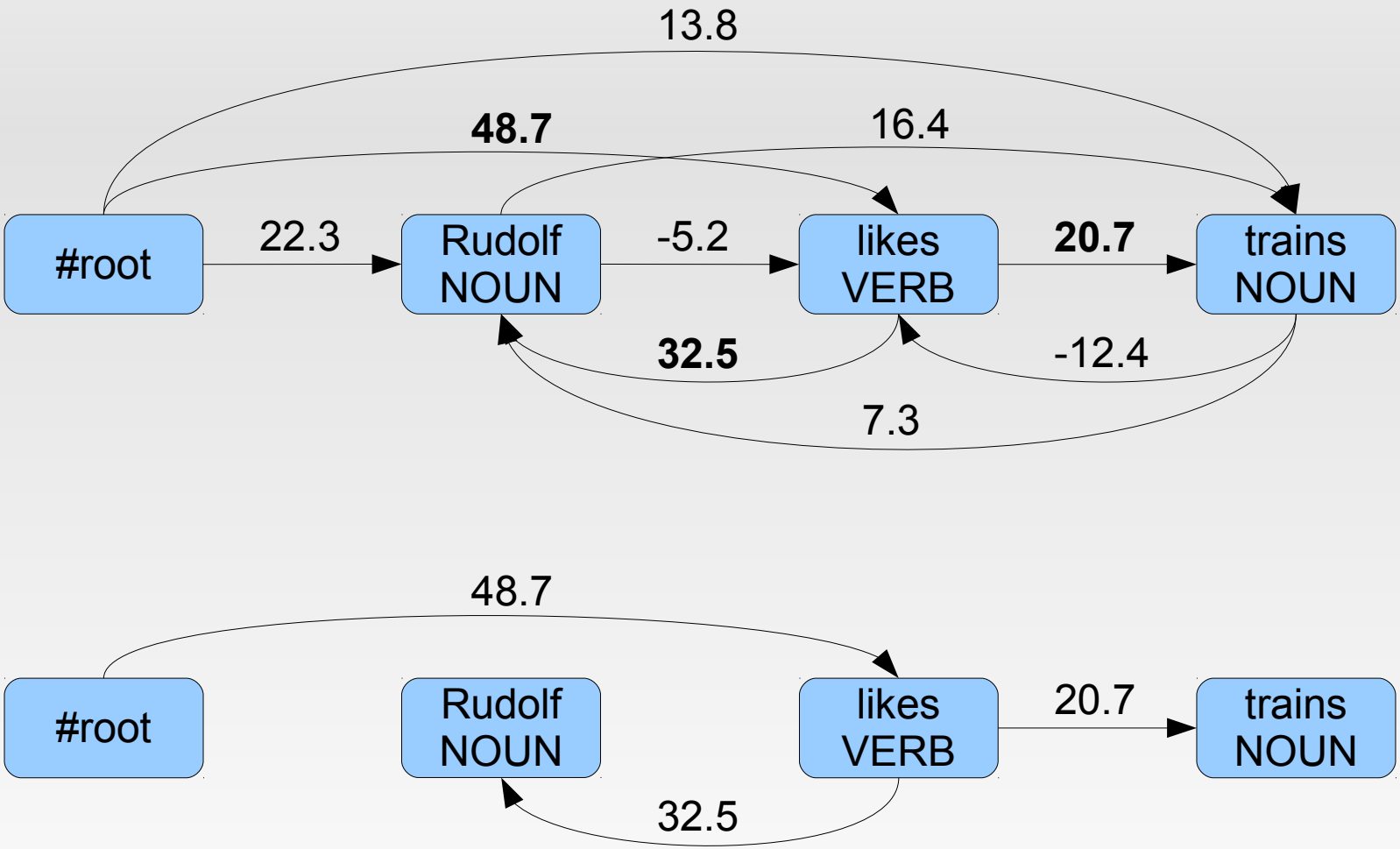
Semi-supervised parsing

- 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 langs (HamleDT: 30 langs)
    - unannotated data (here: POS tagged)

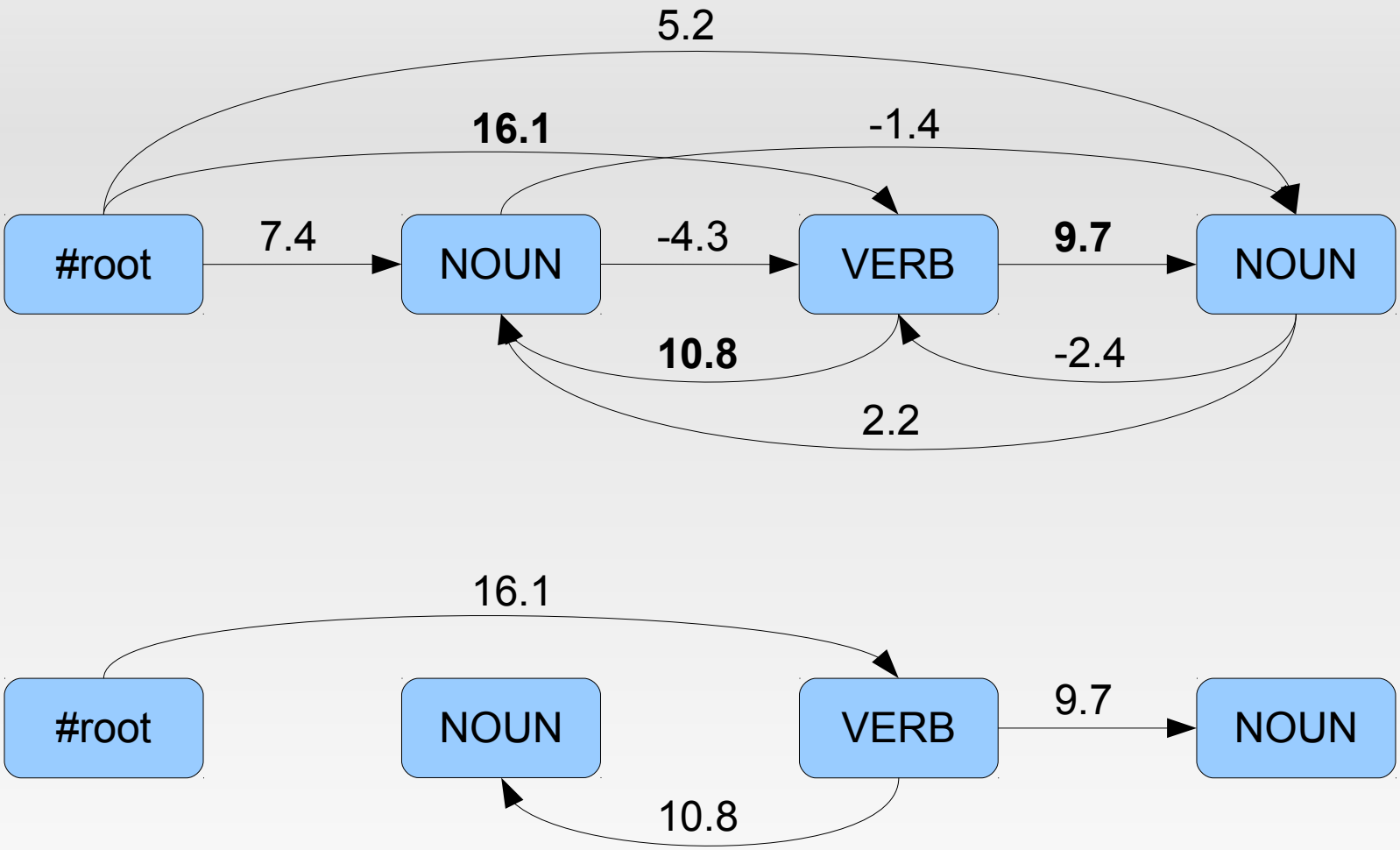
(Lexicalized) MSTParser



(Lexicalized) MSTParser



Delexicalized MSTParser



Single-source delex parser transfer

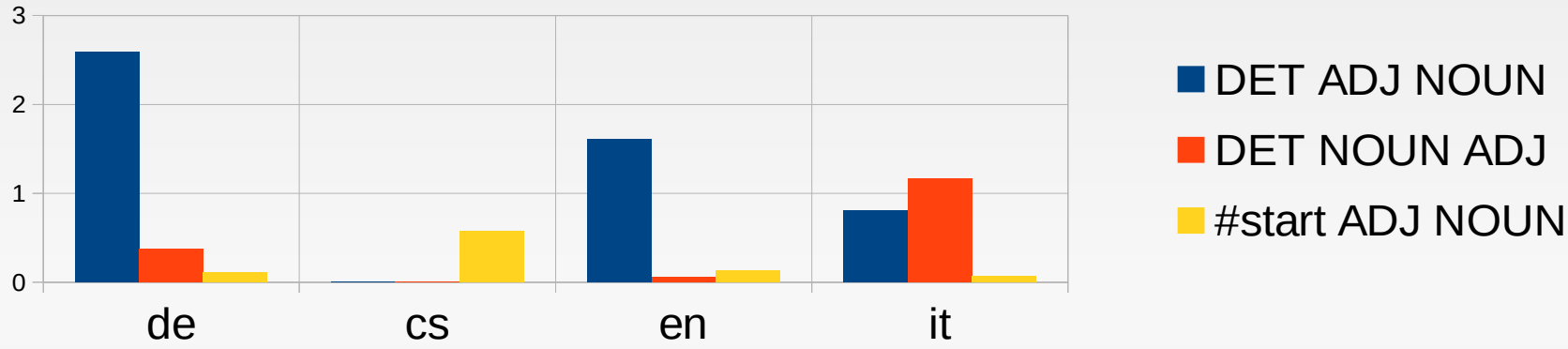
- (Zeman and Resnik, 2008)
- train a delexicalized parser on a source language treebank (e.g. Czech)
- apply it to a target language, without a treebank but with a POS tagger (e.g. Slovak)

Utilizing multiple treebanks

- recall: we have 30 treebanks available
- How do we choose the source treebank?
- Can we use more/all source treebanks?

Choosing the source treebank

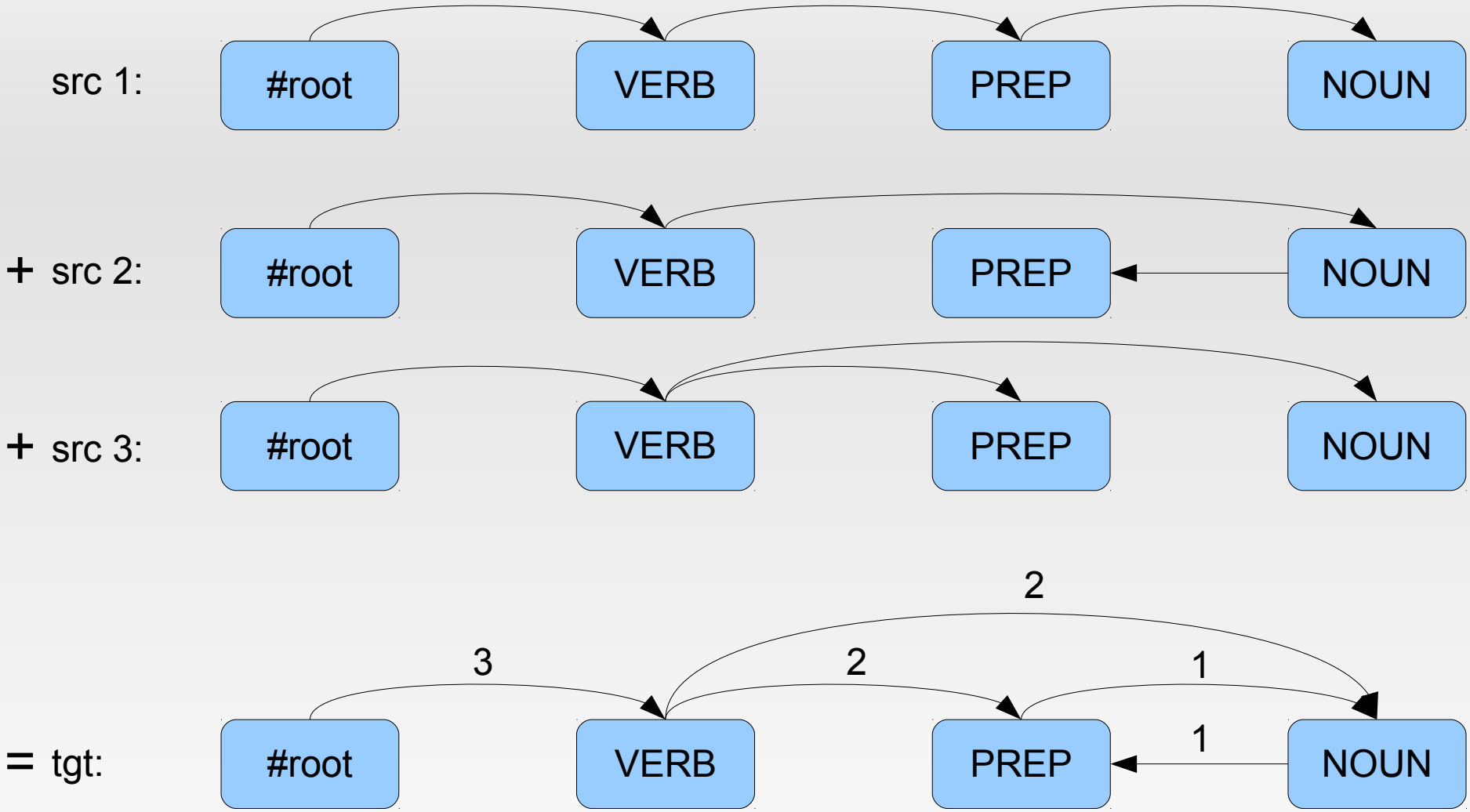
- src should be as similar to tgt as possible
  - WALS (Naseem et al., 2012)
  - POS *n*-gram model (Søgaard and Wulff, 2012)
  - $KL_{cpos3}(tgt,src)$: KL divergence of POS trigram distributions (Rosa and Žabokrtský, 2015)



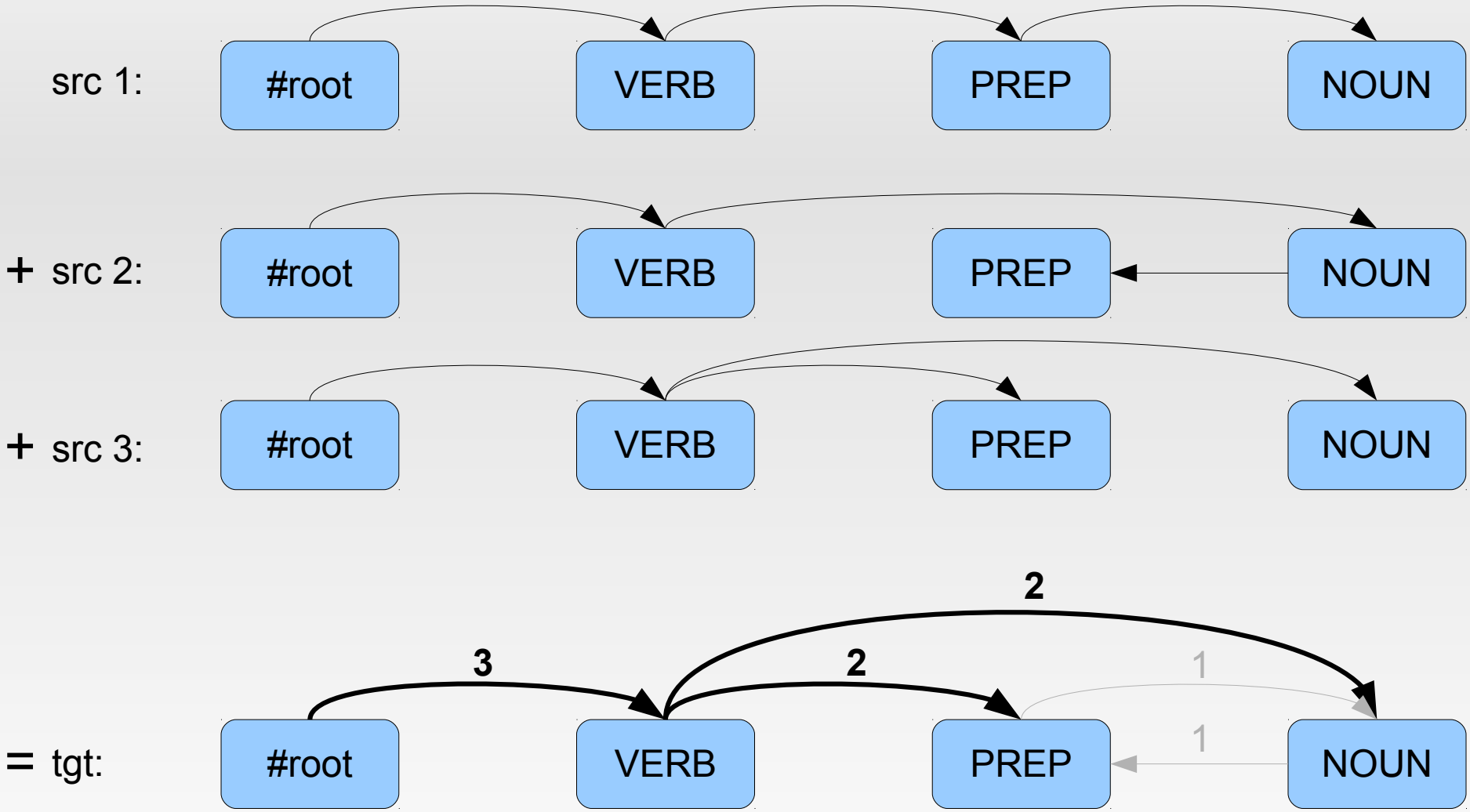
Multi-source delex parser transfer

- treebank concatenation (McDonald et al., 2011)
  - weighted by lang. sim. (Søgaard and Wulff, 2012)
- parse tree combination (Sagae and Lavie, 2006)
  - crosslingual transfer (Rosa and Žabokrtský, 2015)
- parser model interpolation (this work)

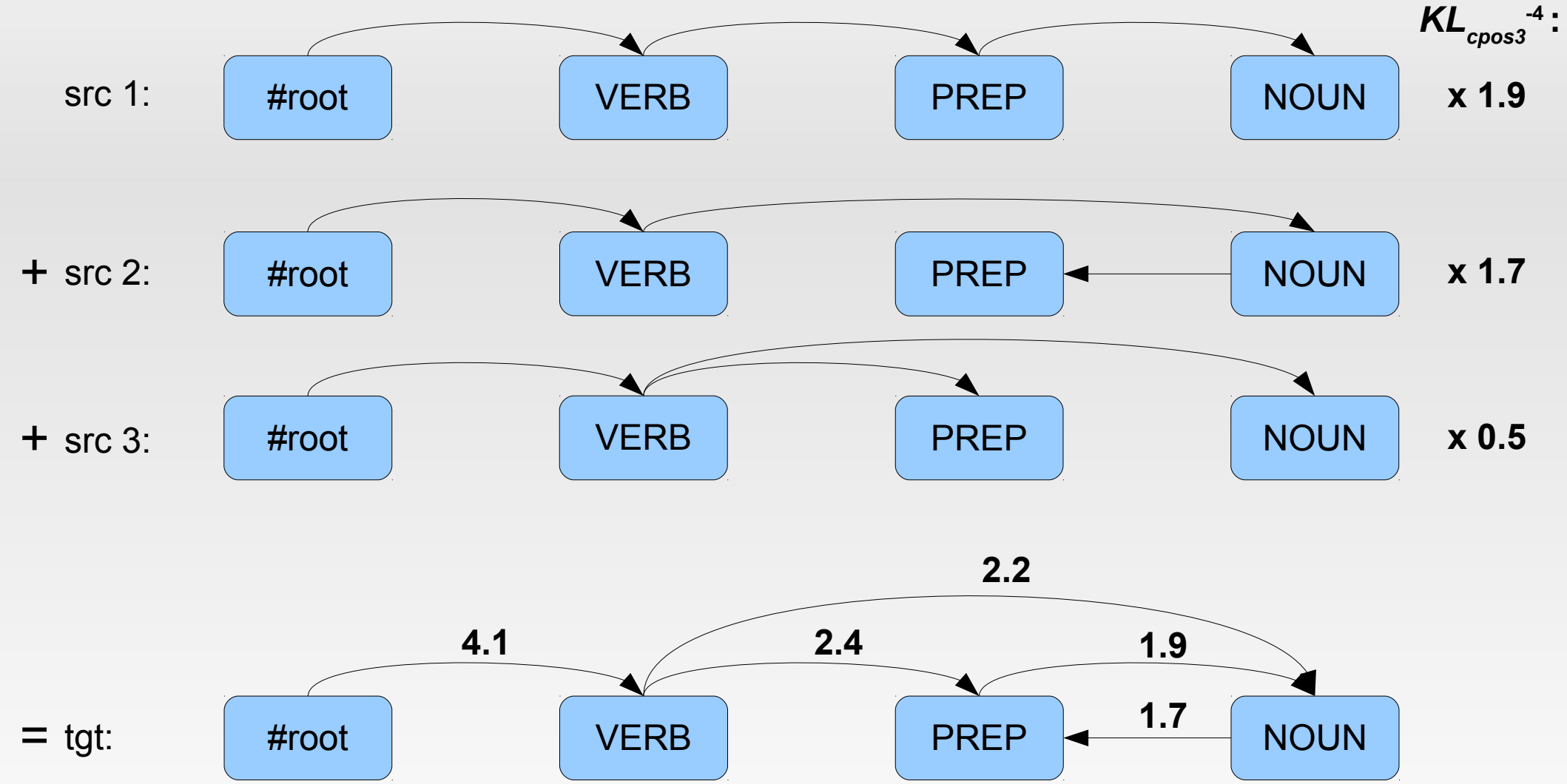
Parse tree combination



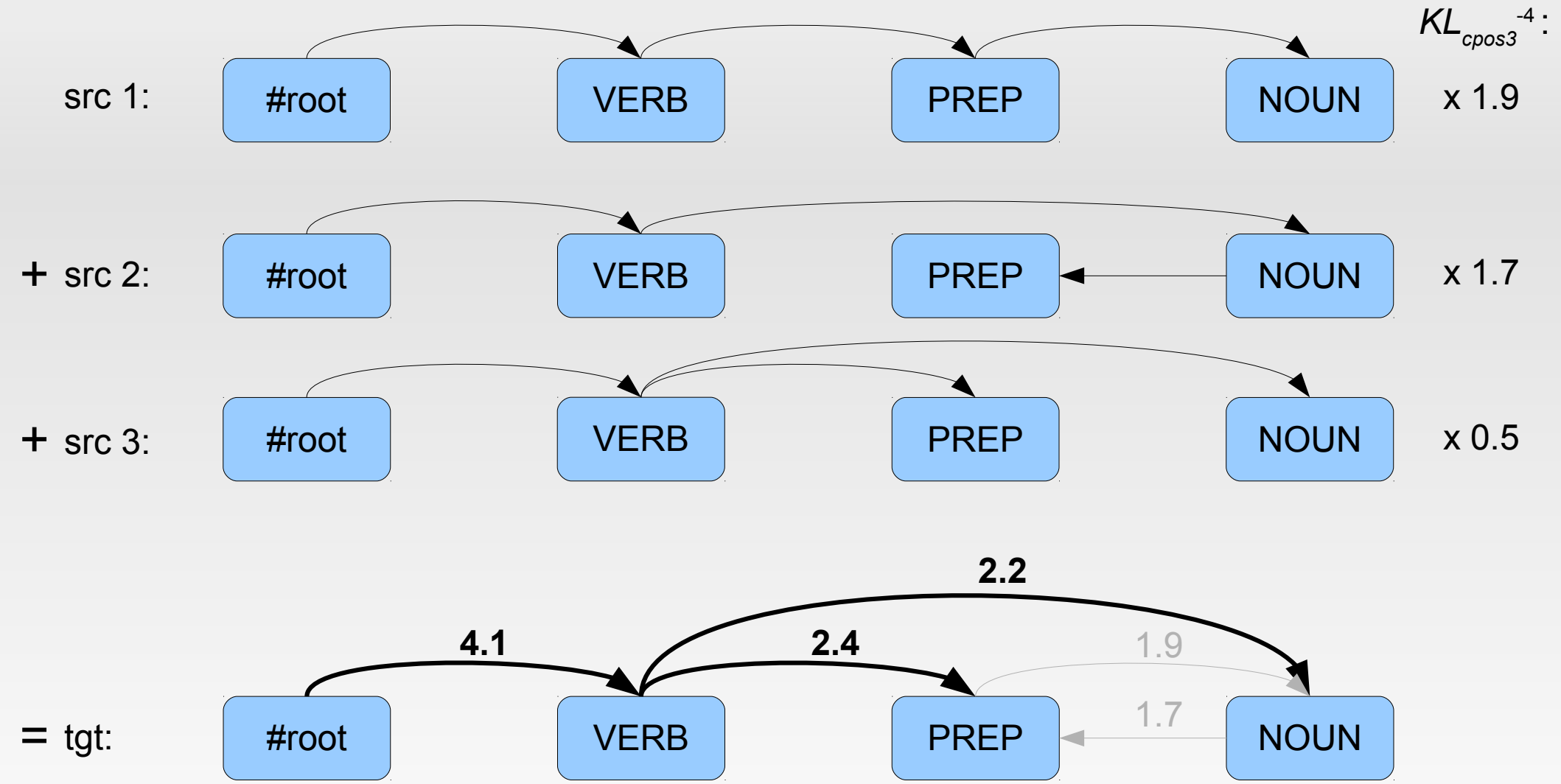
Parse tree combination



Weighted parse tree combination

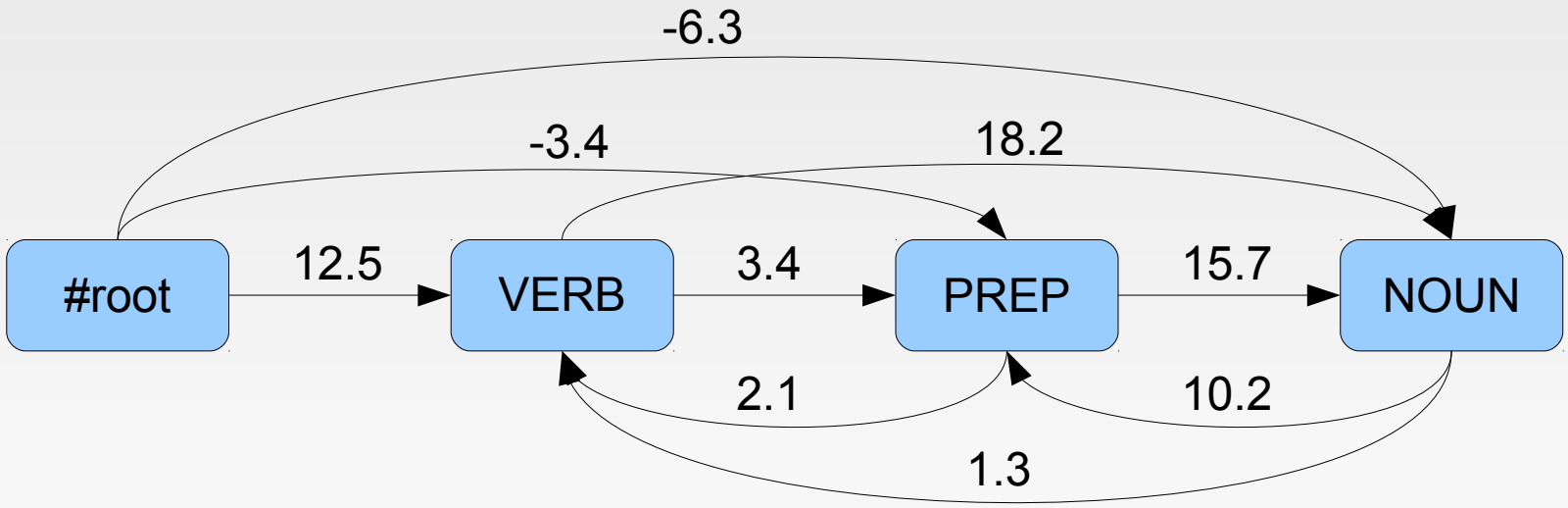


Weighted parse tree combination



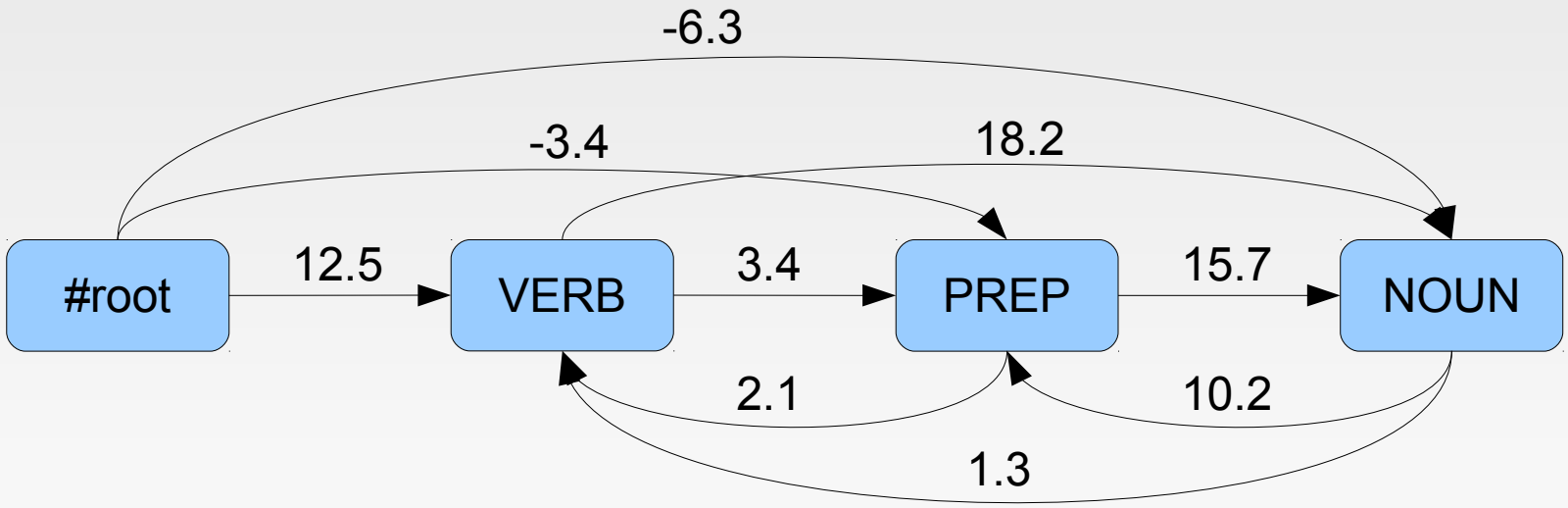
Parser model interpolation

- 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:

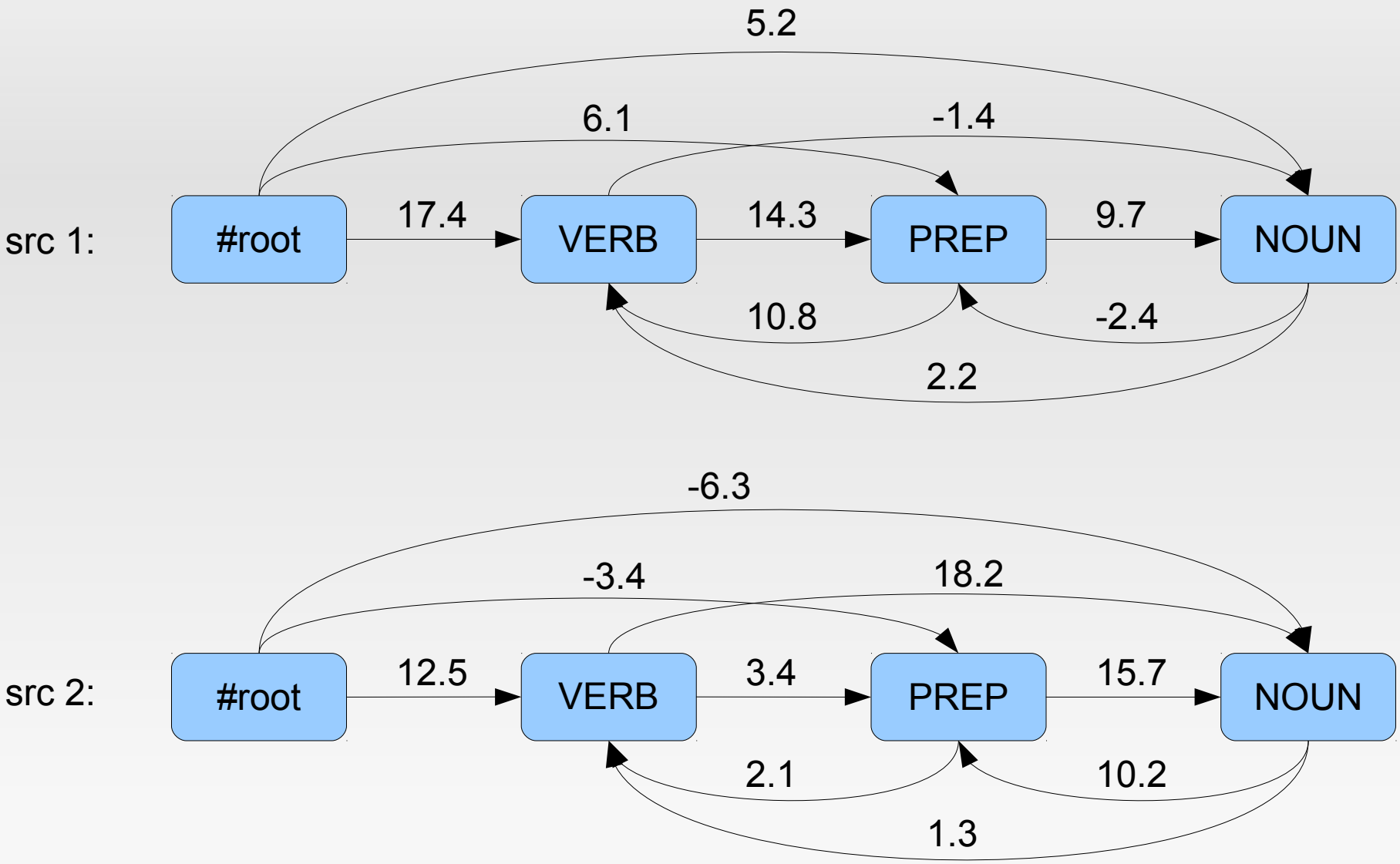


Parser model interpolation

- 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:



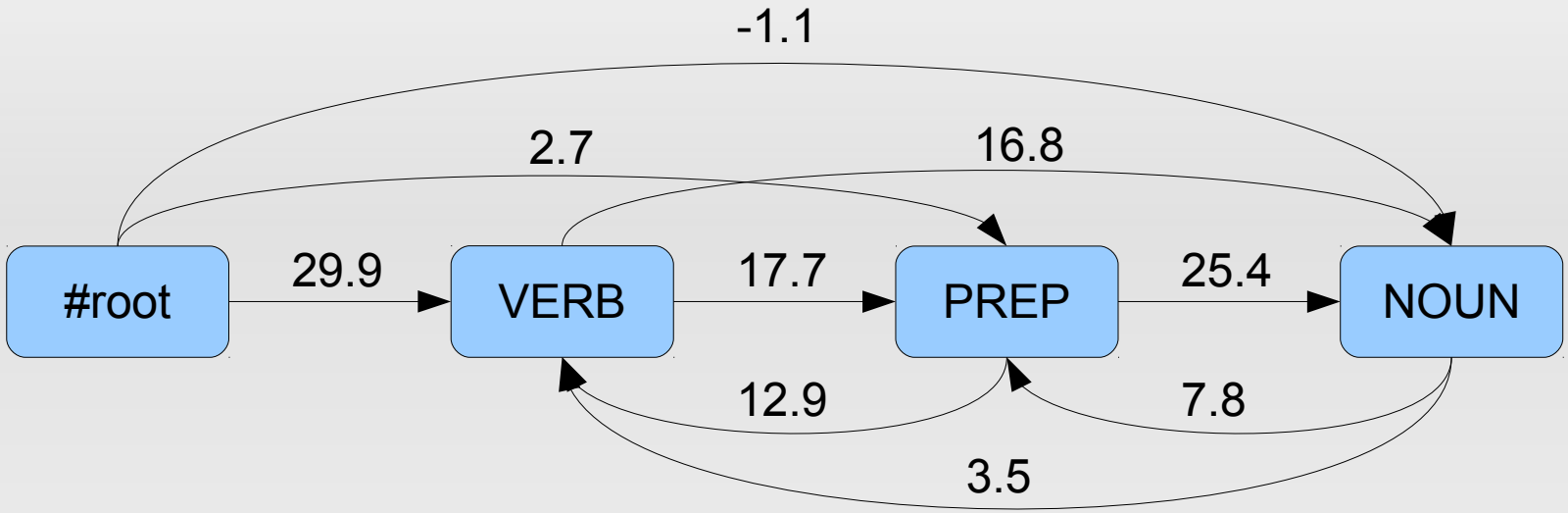
Parser model interpolation



- score normalization!

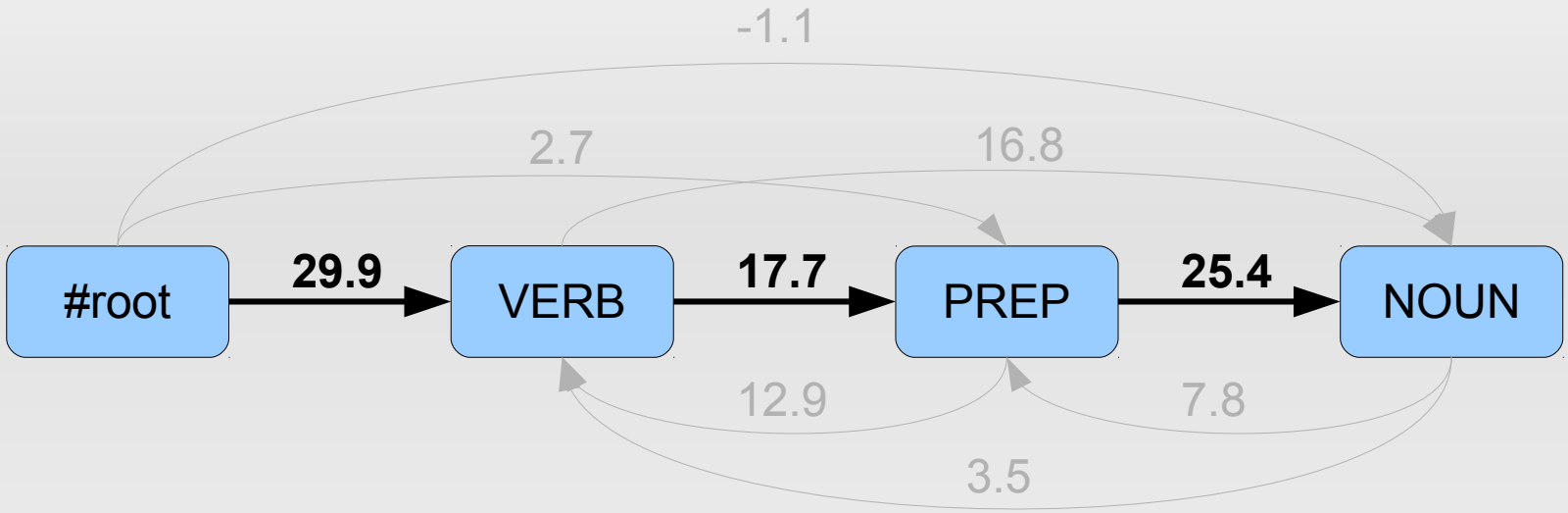
Parser model interpolation

= **tgt (∑):**



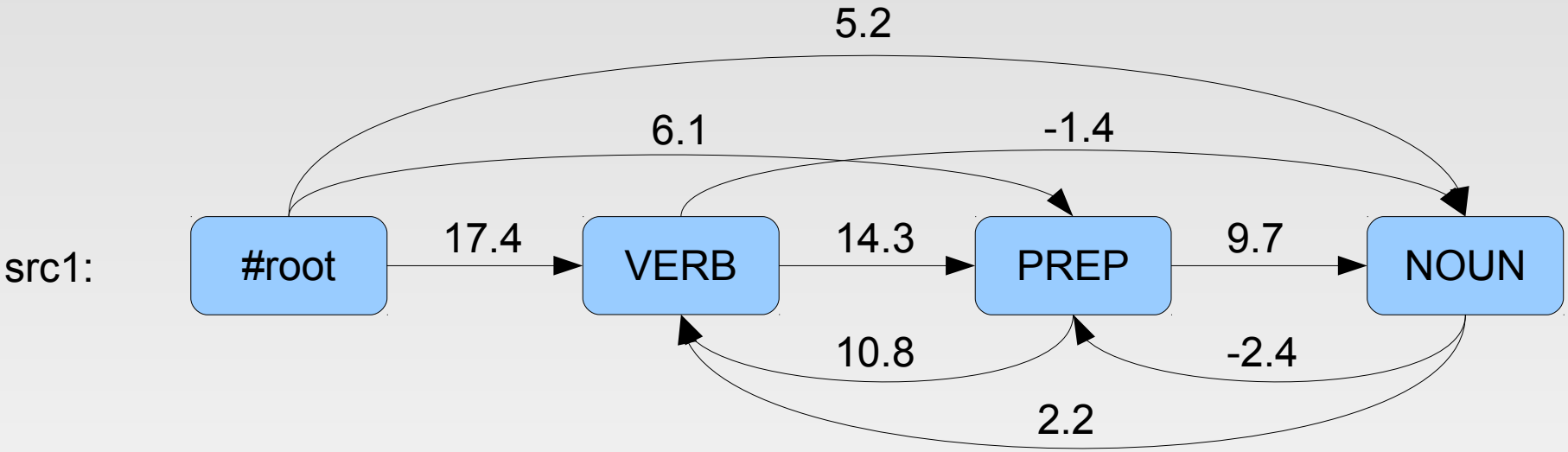
Parser model interpolation

= tgt:



Weighted parser model interpol.

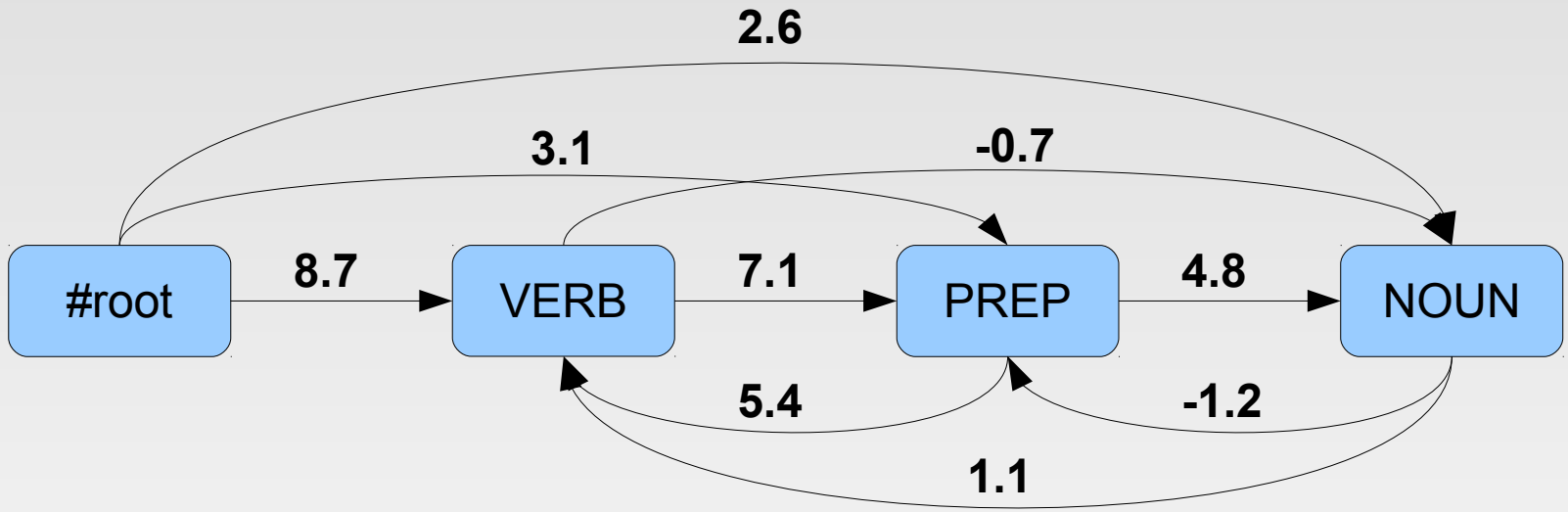
- multiply each edge score with $KL_{cpos3}^{-4}(tgt,src)$



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

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

Why “model interpolation”?

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

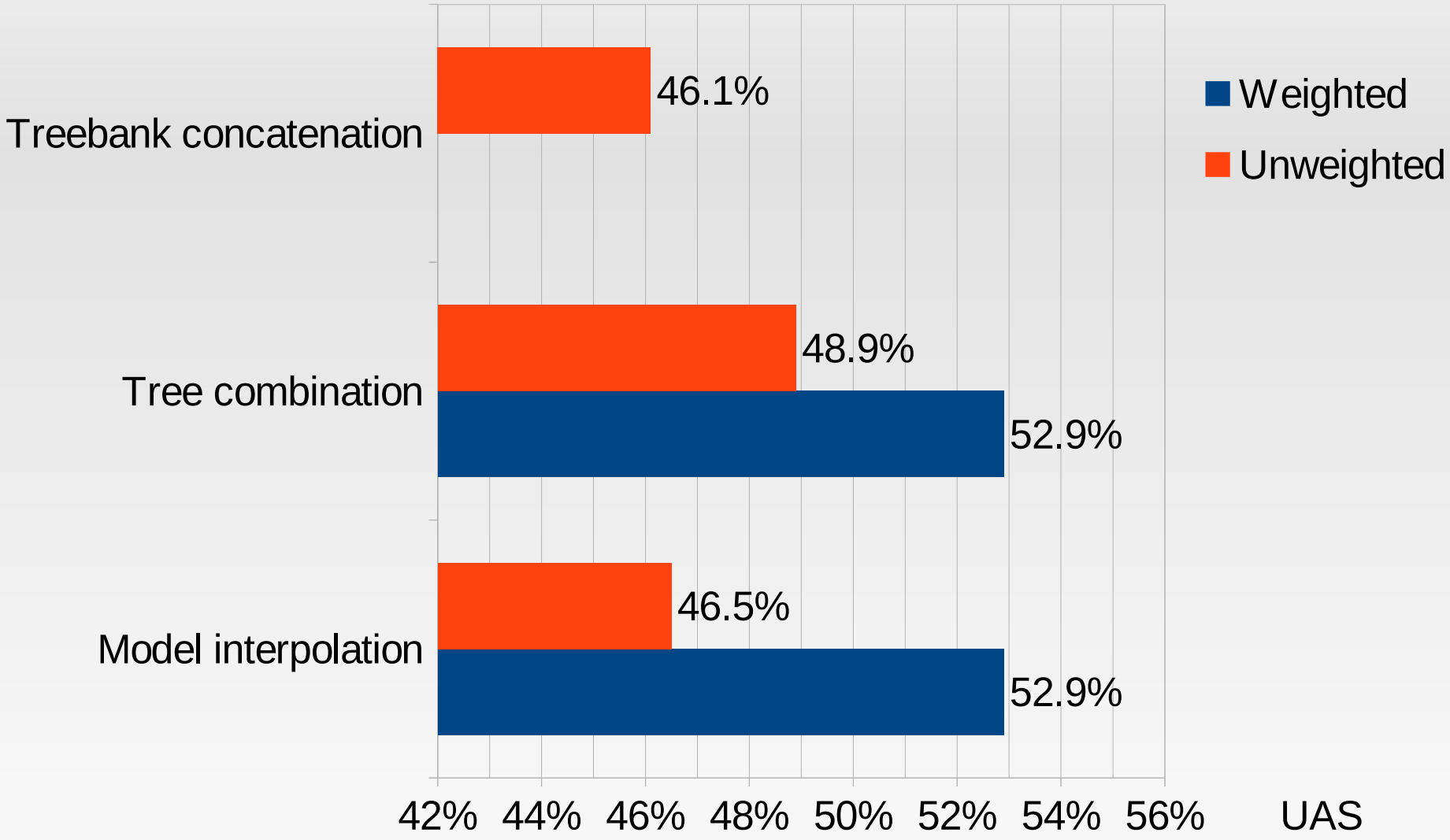
Why “model interpolation”?

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

Why “model interpolation”?

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

Evaluation on HamleDT (30 langs)



Conclusion

- Multi-source delexicalized parser transfer
  - parse tree combination
  - MSTParser model interpolation
  - KL_{cpos3}: language similarity for src selection/weighting
- Weighted model interpolation
  - similar accuracy to tree combination
  - faster inference
  - edge score not a good indicator of parser confidence; better methods exist (Mejer and Crammer, 2012)

Thank you for your attention

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

MSTParser Model Interpolation for Multi-source Delexicalized Transfer

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

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