Universität Stuttgart

Moritz Stiefel
& Ngoc Thang Vu

Enriching ASR Lattices with POS Tags for Dependency Parsing

Title: Enriching ASR Lattices with POS Tags for Dependency Parsing

Motivation

Parsing speech

- POS tags (or other labels) are helpful to downstream tasks
- Lattice-level tags allow for further task integration

Title: Enriching ASR Lattices with POS Tags for Dependency Parsing

Motivation Parsing speech
- POS tags (or other labels) are helpful to downstream tasks
- Lattice-level tags allow for further task integration

First step POS-enriched ASR word lattices

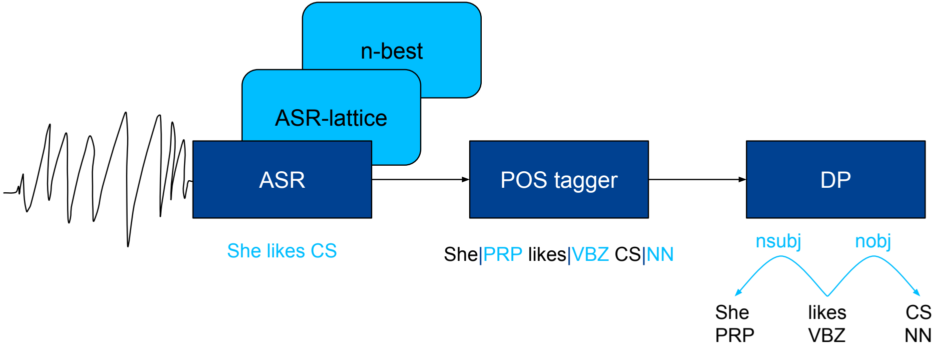
Title: Enriching ASR Lattices with POS Tags for Dependency Parsing

Motivation Parsing speech

- POS tags (or other labels) are helpful to downstream tasks
- Lattice-level tags allow for further task integration

First step POS-enriched ASR word lattices

A pipeline approach:



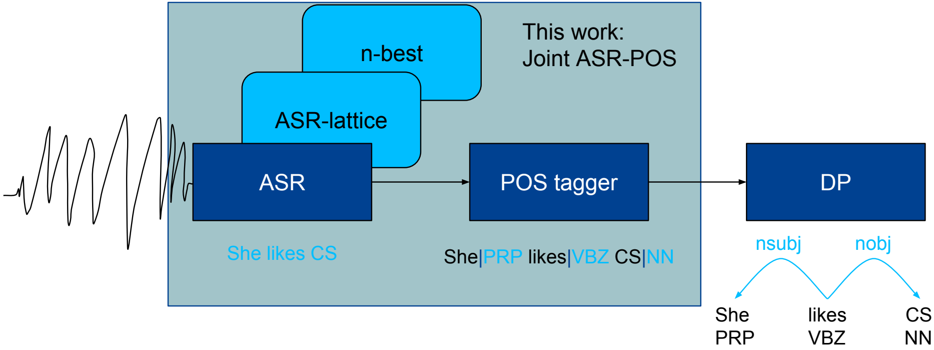
Title: Enriching ASR Lattices with POS Tags for Dependency Parsing

Motivation Parsing speech

- POS tags (or other labels) are helpful to downstream tasks
- Lattice-level tags allow for further task integration

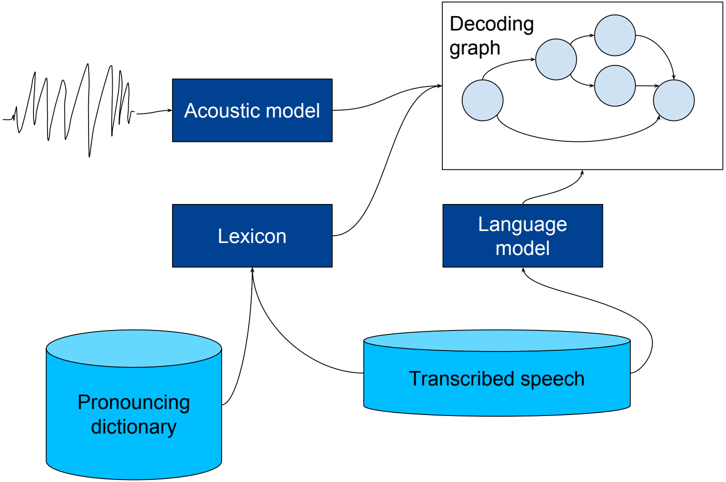
First step POS-enriched ASR word lattices

Our approach:



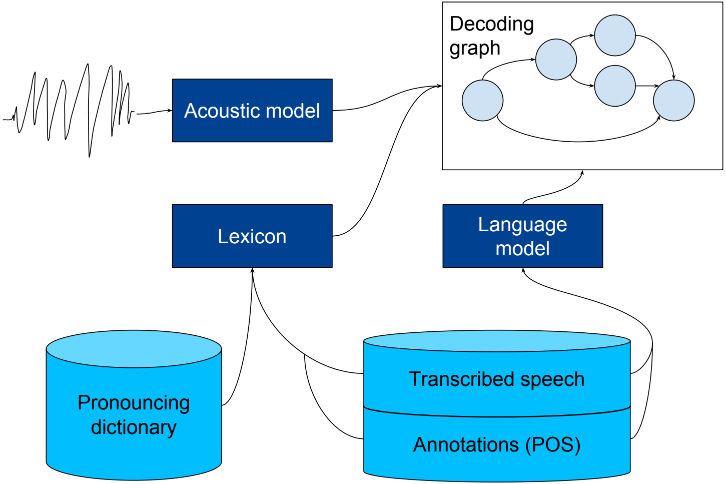
Method

- Using the Kaldi ASR toolkit (Povey et al., 2011)



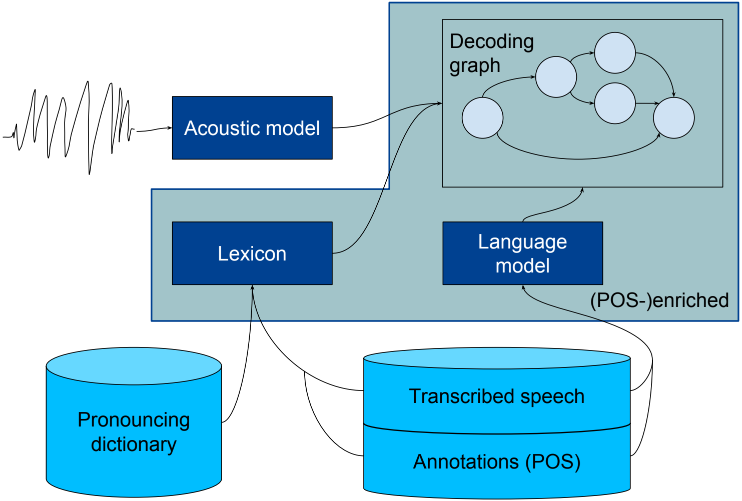
Method

- Using the Kaldi ASR toolkit (Povey et al., 2011)

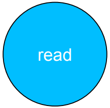


Method

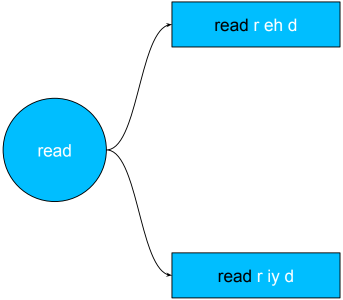
- Using the Kaldi ASR toolkit (Povey et al., 2011)



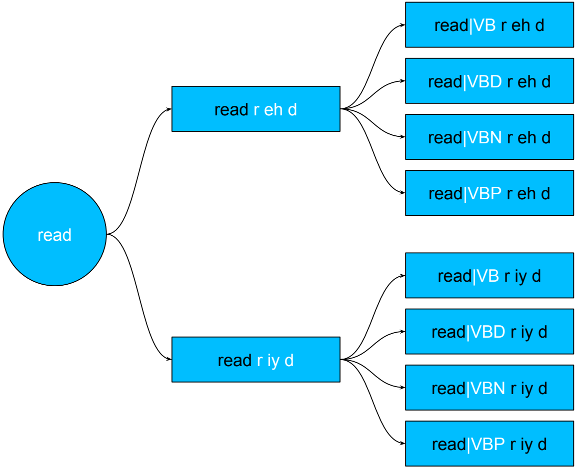
A word-POS paired lexicon



A word-POS paired lexicon



A word-POS paired lexicon



Data: Switchboard splits

- North-American English
- Treebank-3 transcription (not MS-State transcription!)

Set	Conversations	Utterances	Tokens	Avg. tok./utt.	Vocabulary
train	2xxx-3xxx	90823	677160	7.46	14759
dev	4519-4936	5697	50148	8.80	3761
eval	4004-4153	5822	48320	8.30	3695
lmdev	4154-4483	5949	50017	8.41	3742

Table: Summary of SWBD data splits

Data: Switchboard splits

- North-American English
- Treebank-3 transcription (not MS-State transcription!)

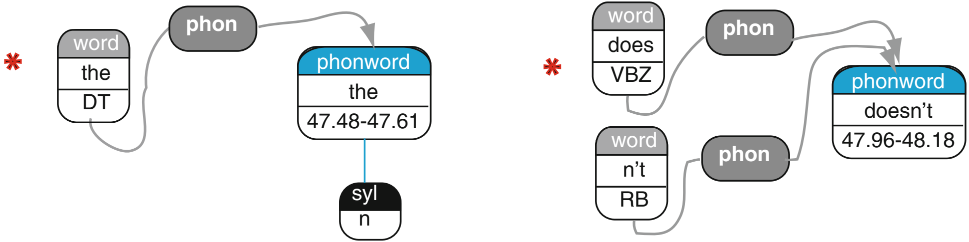
Set	Conversations	Utterances	Tokens	Avg. tok./utt.	Vocabulary
train	2xxx-3xxx	90823	677160	7.46	14759
dev	4519-4936	5697	50148	8.80	3761
eval	4004-4153	5822	48320	8.30	3695
lmdev	4154-4483	5949	50017	8.41	3742

Table: Summary of SWBD data splits

LM	Baseline 2-gram	Baseline 3-gram	Joint 2-gram	Joint 3-gram
PPL	89.4	76.3	96.4	84.2

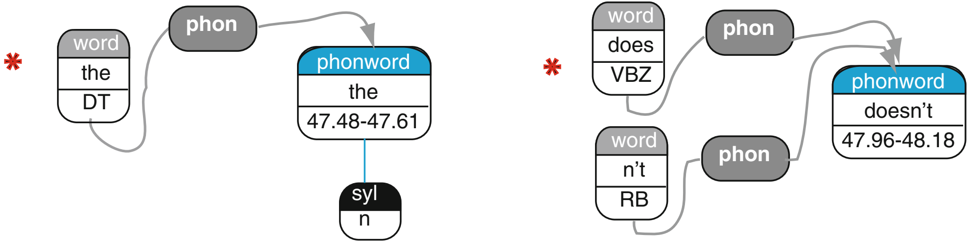
Table: Language model (LM) perplexities (PPL) on *lmdev*.

Data: Switchboard POS-enriched transcription



MS-State vs Treebank-3 transcription, from Calhoun et al. (2010, p. 392)

Data: Switchboard POS-enriched transcription

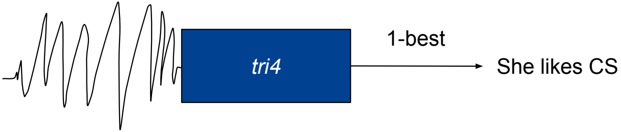


- Orthography/tokenization and POS tags from the Treebank data (word)
- Timestamps from linked MS-State transcriptions (phonword)

MS-State vs Treebank-3 transcription, from Calhoun et al. (2010, p. 392)

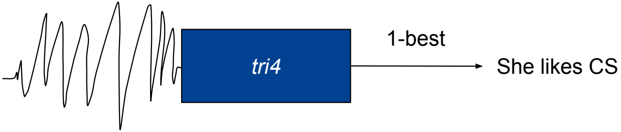
Intermediate results: ASR

tri4

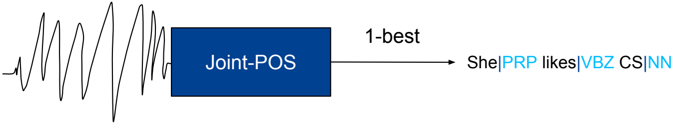


Intermediate results: ASR

tri4

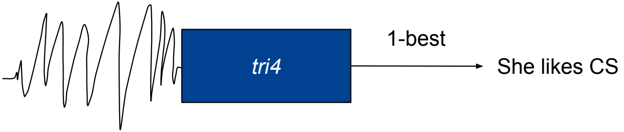


Joint-POS

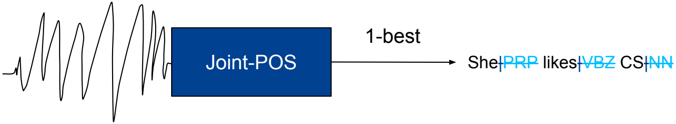


Intermediate results: ASR

tri4



Joint-POS

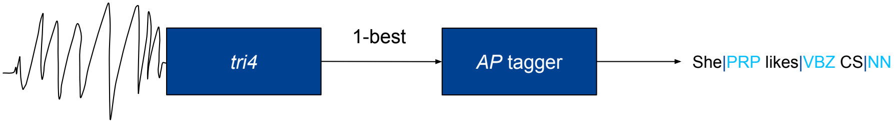


Set	*tri4*	Joint-POS
dev	**28.75** (65.83)	28.93 **(65.28)**
test	29.41 (64.41)	**29.26 (64.15)**

Table: ASR results: numbers are WER (SER)

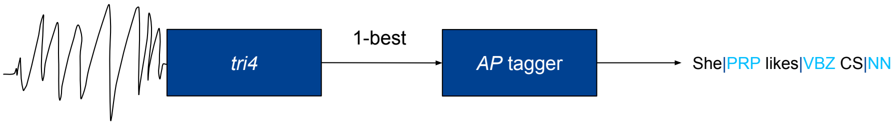
Intermediate results: POS

tri4+AP

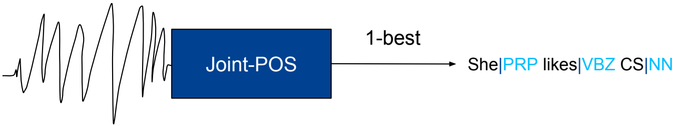


Intermediate results: POS

tri4+AP

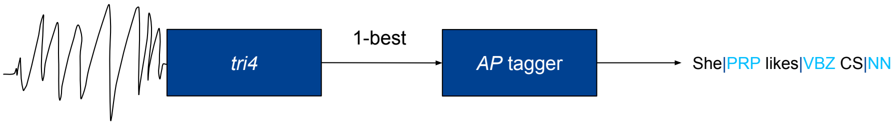


Joint-POS

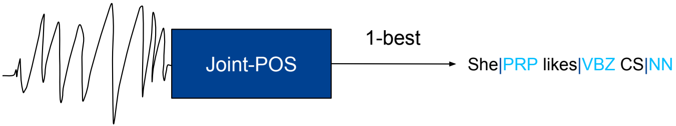


Intermediate results: POS

tri4+AP



Joint-POS



Set	*tri4+ME.pre*	*tri4+AP.pre*	*tri4+spaCy.pre*	*tri4+ME.70k*	*tri4+AP*	Joint-POS
dev	43.29 (94.23)	45.46 (95.84)	39.17 (82.38)	33.24 (68.18)	32.30 (67.67)	**32.05 (67.32)**
test	44.49 (94.19)	46.18 (95.74)	40.42 (81.86)	36.23 (67.26)	33.10 (66.85)	**32.52 (66.52)**

Table: POS tagging results: numbers are WER (SER)

DP results

- 1-best hypotheses of standard Kaldi *tri4* setup plus AP tagger vs our Joint-POS
- Xiang Yu's parser after (Weiss et al., 2015):
  greedy neural transition-based parser, uses word and POS features

DP results

- 1-best hypotheses of standard Kaldi *tri4* setup plus AP tagger vs our Joint-POS
- Xiang Yu's parser after (Weiss et al., 2015):
  greedy neural transition-based parser, uses word and POS features

			tri4+AP		Joint-POS	
Set	#utts	#tokens	UAS	LAS	UAS	LAS
dev	900	4881	94.30	92.71	**95.41**	**93.63**
test	882	4827	94.68	93.06	**94.92**	**93.52**

Table: Parsing results for subsets of correct tokenizations. Labeled attachment scores (LAS) and unlabeled attachment scores (UAS) given as percentages.

DP results

- 1-best hypotheses of standard Kaldi *tri4* setup plus AP tagger vs our Joint-POS
- Xiang Yu's parser after (Weiss et al., 2015):
  greedy neural transition-based parser, uses word and POS features

			tri4+AP		Joint-POS	
Set	#utts	#tokens	UAS	LAS	UAS	LAS
dev	900	4881	94.30	92.71	**95.41**	**93.63**
test	882	4827	94.68	93.06	**94.92**	**93.52**

Table: Parsing results for subsets of correct tokenizations. Labeled attachment scores (LAS) and unlabeled attachment scores (UAS) given as percentages.

- High scores, but only on utterances with **correct** tokenizations

DP results extended

- Number of correctly tokenized utterances $\neq$ number of utterances

DP results extended

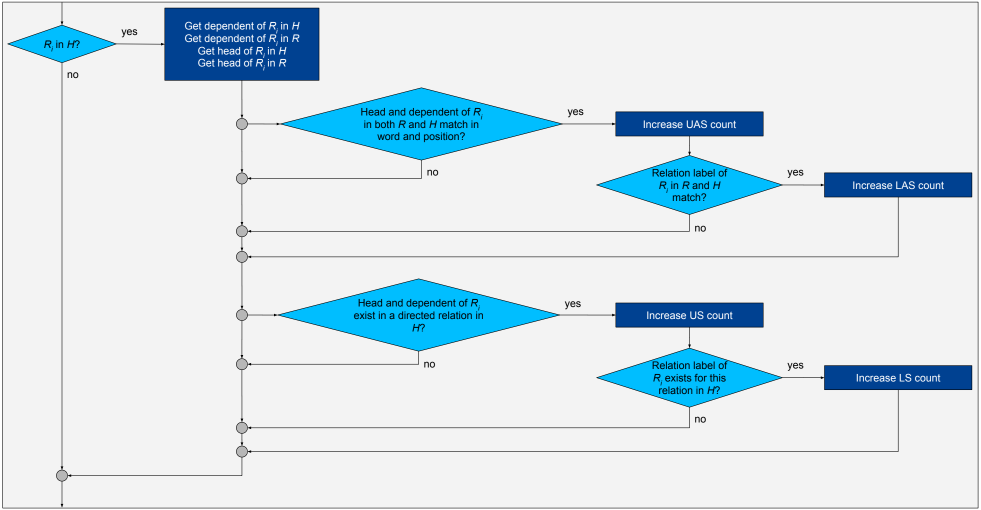
- Number of correctly tokenized utterances $\neq$ number of utterances
  - How can we evaluate incorrectly recognized utterances?

DP results extended

- Number of correctly tokenized utterances $\neq$ number of utterances
  - How can we evaluate incorrectly recognized utterances?
  - Our answer: fuzzy relation-based measure that ignores word position altogether

DP results extended: fuzzy relation-based measure for US and LS

Initialize UAS, LAS, US and LS with zero count
For all reference utterances *R* that have a hypothesis *H*
For all tokens R_i in *R*



DP results extended

- Number of correctly tokenized utterances $\neq$ number of utterances
  - How can we evaluate incorrectly recognized utterances?
  - Our answer: fuzzy relation-based measure that ignores word position altogether

DP results extended

- Number of correctly tokenized utterances $\neq$ number of utterances
  - How can we evaluate incorrectly recognized utterances?
  - Our answer: fuzzy relation-based measure that ignores word position altogether

Model	Set	UAS	LAS	US	LS
tri4+AP	dev	32.20	31.20	52.02	49.40
	test	31.21	30.29	50.72	48.33
Joint-POS	dev	**32.41**	**31.43**	**52.21**	**49.71**
	test	**31.56**	**30.73**	**51.21**	**48.99**

Table: Parsing results on full *dev* and *test* sets. LAS and UAS given as percentages. LS (labeled score) and US (unlabeled score) are a fuzzy evaluation metric devised to be able to evaluate tokenization mismatches between the ASR hypotheses and the corresponding treebank data. LS and US are also given as percentages. The *dev* set has 3994 utterances with 44760 tokens and the *test* set has 3912 utterances with 43277 tokens. Best scores per set are bold-faced.

DP-based error analysis 1/3

- *tri4* token incorrect, subsequent POS tag, too

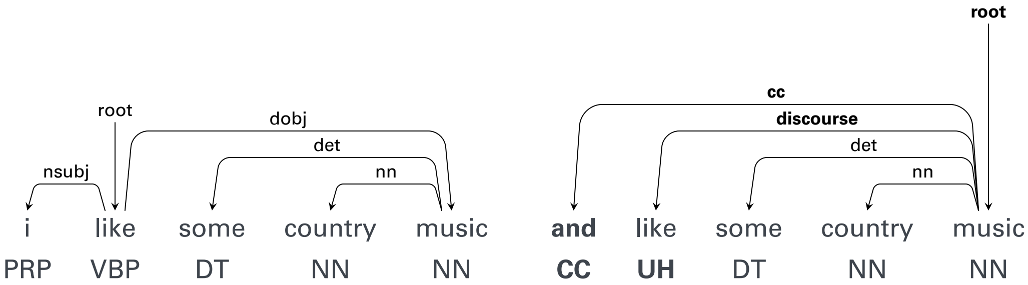


Figure: Correct Joint-POS tree on the left, incorrect *tri4* tree on the right.

DP-based error analysis 2/3

- *tri4* ASR deletion error

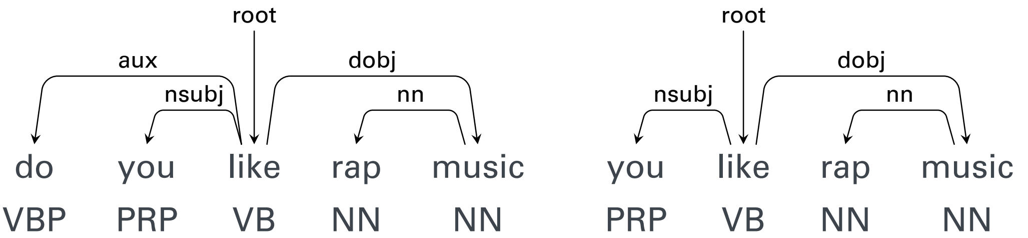


Figure: Correct Joint-POS tree on the left, incorrect *tri4* tree on the right.

DP-based error analysis 3/3

- Joint-POS token sequence incorrect resulting in erroneous parse

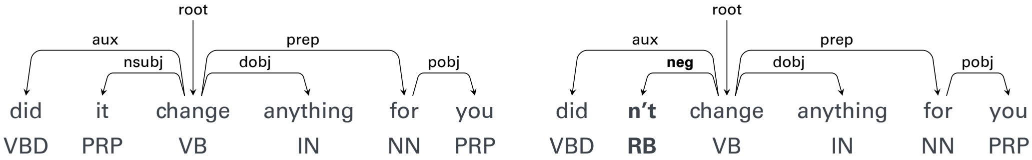


Figure: Correct *tri4* tree on the left, incorrect Joint-POS tree on the right.

Conclusions

- Successful joint ASR and POS tagging
  - Increased search space in the decoding graph
  - No performance loss compared to pipeline approach

⇒ **POS tags in an ASR lattice structure**

- Possible avenues of exploration:
  - Systematic error analysis
  - Use transcriptions tagged with a POS-tagger and compare results
  - Comparison against the approach of Velikovich (2016), who tag lattices

Universität Stuttgart

Moritz Stiefel & Ngoc Thang Vu
Institut für Maschinelle Sprachverarbeitung (IMS), Universität Stuttgart

eMail moritz.stiefel@ims.uni-stuttgart.de
Telefon +49-711-685 813 60
Fax +49-711-685 813 66

References

Sasha Calhoun, Jean Carletta, Jason M. Brenier, Neil Mayo, Dan Jurafsky, Mark Steedman, and David Beaver. The NXT-format switchboard corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language Resources and Evaluation*, 44(4):387–419, 2010. ISSN 1574-0218. doi: 10.1007/s10579-010-9120-1. URL http://dx.doi.org/10.1007/s10579-010-9120-1.

Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, Jan Silovsky, Georg Stemmer, and Karel Vesely. The kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society, December 2011. IEEE Catalog No.: CFP11SRW-USB.

Leonid Velikovich. Semantic model for fast tagging of word lattices. In *2016 IEEE Spoken Language Technology Workshop, SLT 2016, San Diego, CA, USA, December 13-16, 2016*, pages 398–405. IEEE, 2016. doi: 10.1109/SLT.2016.7846295. URL https://doi.org/10.1109/SLT.2016.7846295.

David Weiss, Chris Alberti, Michael Collins, and Slav Petrov. Structured training for neural network transition-based parsing. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*, pages 323–333. The Association for Computer Linguistics, 2015. ISBN 978-1-941643-72-3. URL http://aclweb.org/anthology/P/P15/P15-1032.pdf.

This work was funded by the German Research Foundation (DFG) through the Collaborative Research Center (SFB) 732, project A8, at the University of Stuttgart.

DP-based error analysis extra 1/2

- *tri4* token incorrect, subsequent POS tag, too

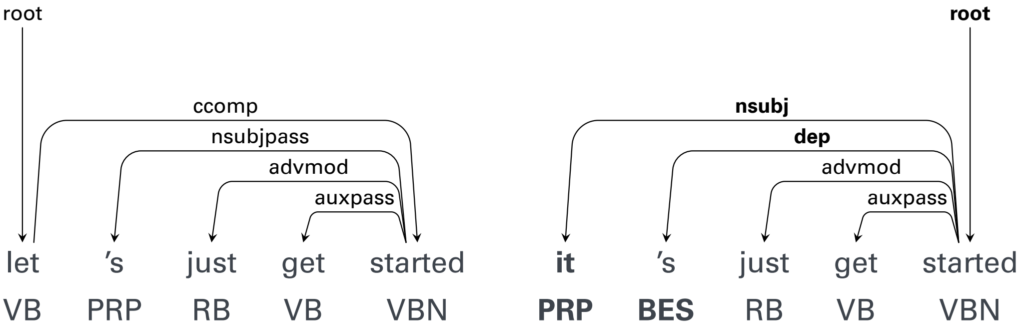


Figure: Correct Joint-POS tree on the left, incorrect *tri4* tree on the right.

DP-based error analysis extra 2/2

- *tri4* with correct tokenization, but POS tagging error

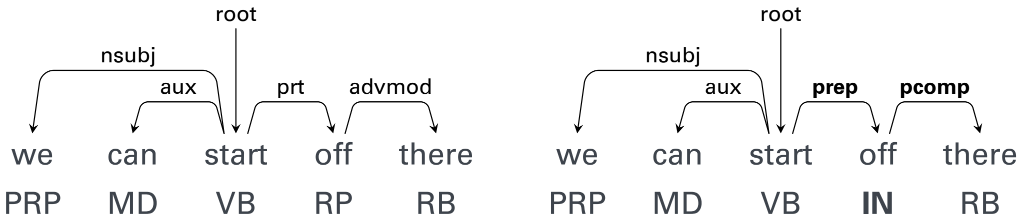


Figure: Correct Joint-POS tree on the left, incorrect *tri4* tree on the right.