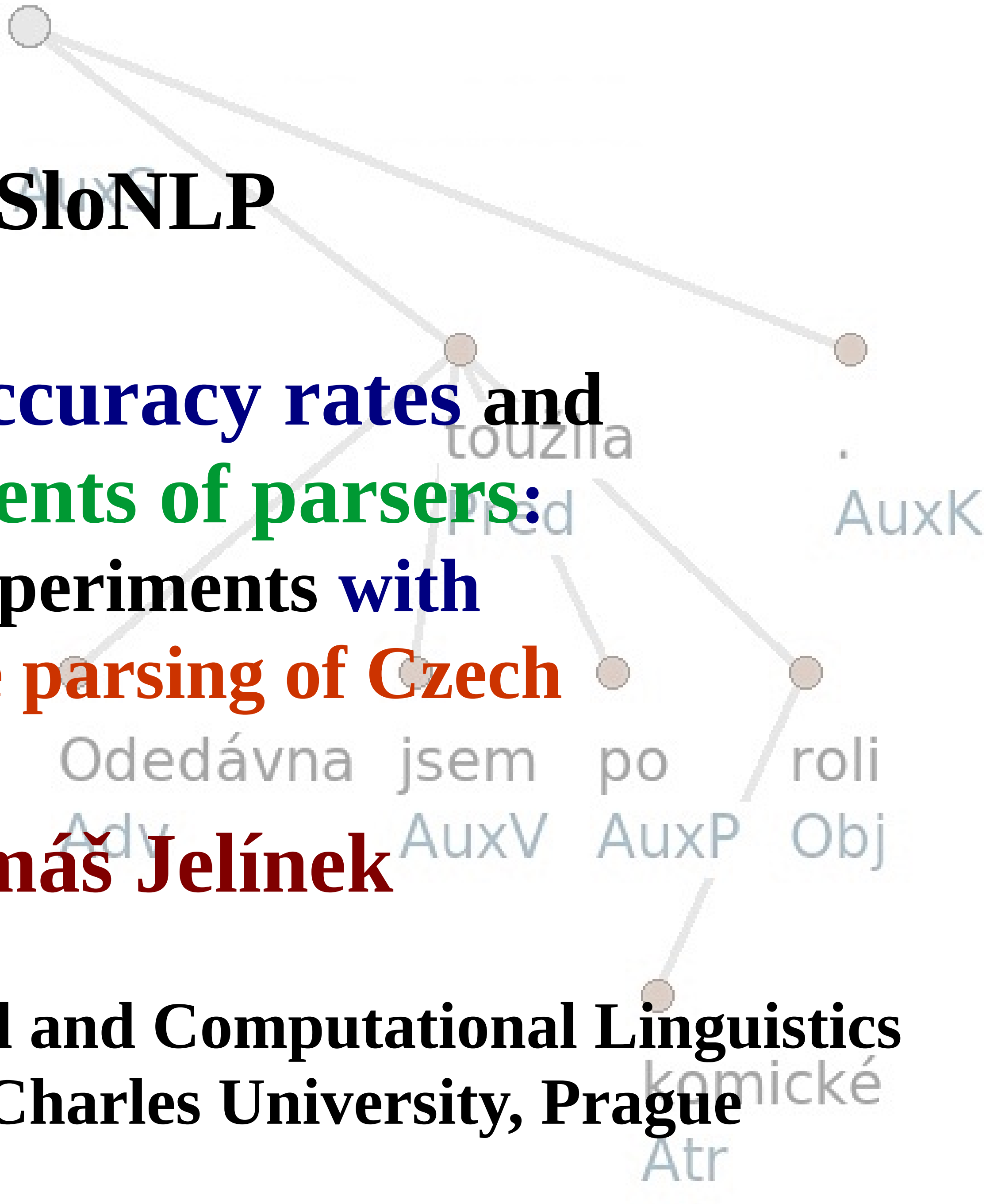
SloNLP

Partial accuracy rates and agreements of parsers: two experiments with ensemble parsing of Czech

Tomáš Jelínek

Institute of Theoretical and Computational Linguistics
Faculty of Arts, Charles University, Prague



Outline

- **Parsers and data**
- **Improve ensemble parsing by using information gathered about the behaviour of the parsers**

- **Partial accuracy rates**
- **Agreements of parsers**

- **Results**

PDT a-layer data (train, dtest, etest)

Dependency Parsers

	UAS	LAS	SENT_U	SENT_L
MST	86.41	79.30	38.93	24.64
Malt	86.74	81.32	42.40	32.68
Turbo	88.63	82.33	44.63	28.75
Mate	88.50	83.03	45.56	33.59
Parsito	86.71	81.42	41.81	32.65

Improve dependency parsing by combining parsers

Extract detailed information about the behaviour of each parser from parsed train data:

train	dtest	etest
train0.1		
train0.2		
train0.3		
train0.4		
...		
parsed data: extract information		
train	dev	test

Combination of parsers: floor and ceiling

Floor: all parsers correct → correct
Ceiling: any parser correct → correct

	UAS	LAS	SENT_U	SENT_L
Floor	75.76	67.02	25.95	16.84
Mate	88.50	83.03	45.56	33.59
Ceiling	95.72	92.55	69.03	55.27

Partial accuracy rates: features

Accuracy rate: proportion of correct tokens with a given property, for example *nouns* or *nouns dependent on adjectives* (0-1)

Accuracy rates calculated for:

POS	POS of dependent token
2POS	POS of governing token
DIST	distance dep. - gov. tokens (intervals)
POS2POS (P2P)	POS of both
POSCASE	POS + case of dep. token
POSSUBPOS, POSSUBPOSCASE...	

LAS/UAS (labeled / unlabeled)

Nathan D. Green: Improvements to Syntax-based Machine Translation using Ensemble Dependency Parsers

Accuracy rates: example of table entries

MALT	DIST	LAS	2	0.81527
MATE	DIST	LAS	2	0.82592
TURBO	DIST	LAS	2	0.82854
MALT	P2P	UAS	AN	0.96789
MATE	P2P	UAS	AN	0.97104
TURBO	P2P	UAS	AN	0.97101
MALT	P2P	UAS	NR	0.97581
MATE	P2P	UAS	NR	0.97375
TURBO	P2P	UAS	NR	0.97887
MALT	POS	LAS	N	0.79647
MATE	POS	LAS	N	0.80581
TURBO	POS	LAS	N	0.79013

Combining dependency parsers using accuracy rates

1. Calculate accuracy rates ✓
2. Parse new text with all parsers ✓
3. Merge parsed texts ✓
4. Use pre-calculated **best options** for the combination of parsers (hundreds of experiments with features & exponents were run)
   a. Features (POS, DIST etc., LAS/UAS)
   b. Exponent ($(\text{acc. rate})^{exp}$, e.g. 0.79^4,1–9)
5. Calculate weights for each edge (token) using accuracy rates
6. Best syntactic structure is calculated by MST algorithm (a new tree may be created)

Weights calculation

for each edge, given a set of **features** (POS, DIST...) and **exponent** (1 ... 9):

weight of the edge is calculated as the sum of error rates for each **feature**, raised to the power of the **exponent**; if two or more parsers agree on an edge, it is then the sum of such error rates for each parser

Example: features: **POS:LAS**, exponent **2**

Privatizované mlékárny se však zatím ***mezi*** *sebou* ***nedokázaly*** ***domluvit****.*

However, the privatized dairies were not yet able to agree among themselves.

mezi: Mate & Malt: ***nedokázaly***; POS R: 0.829^2+0.809^2=**1.342**

Turbo & MST: ***domluvit***; POS R: 0.832^2+0.814^2=**1.355**

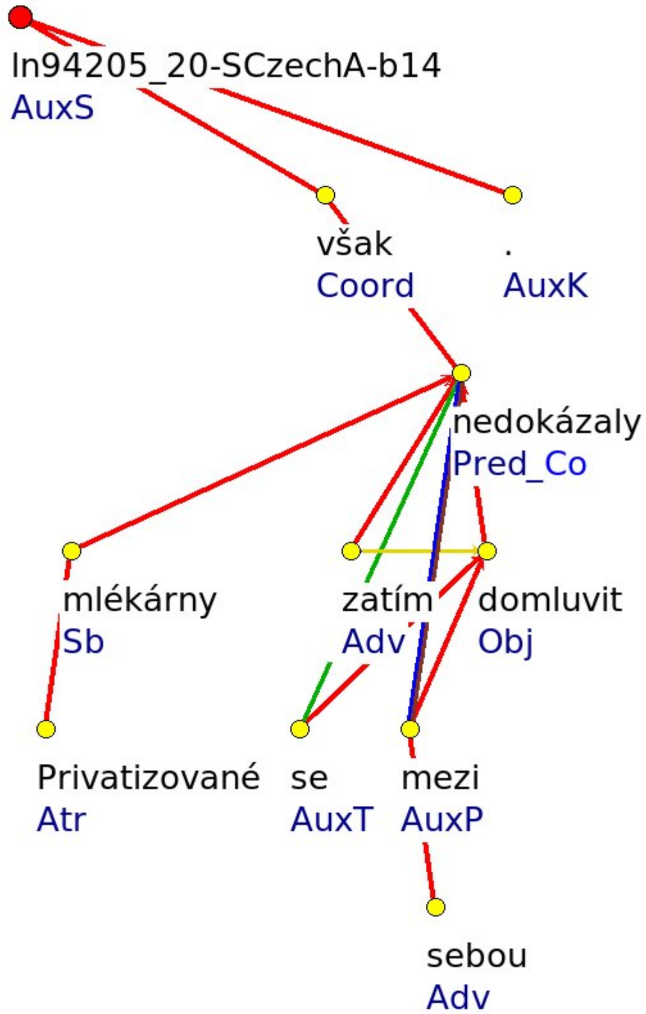
MST algorithm using weights based on accuracy rates

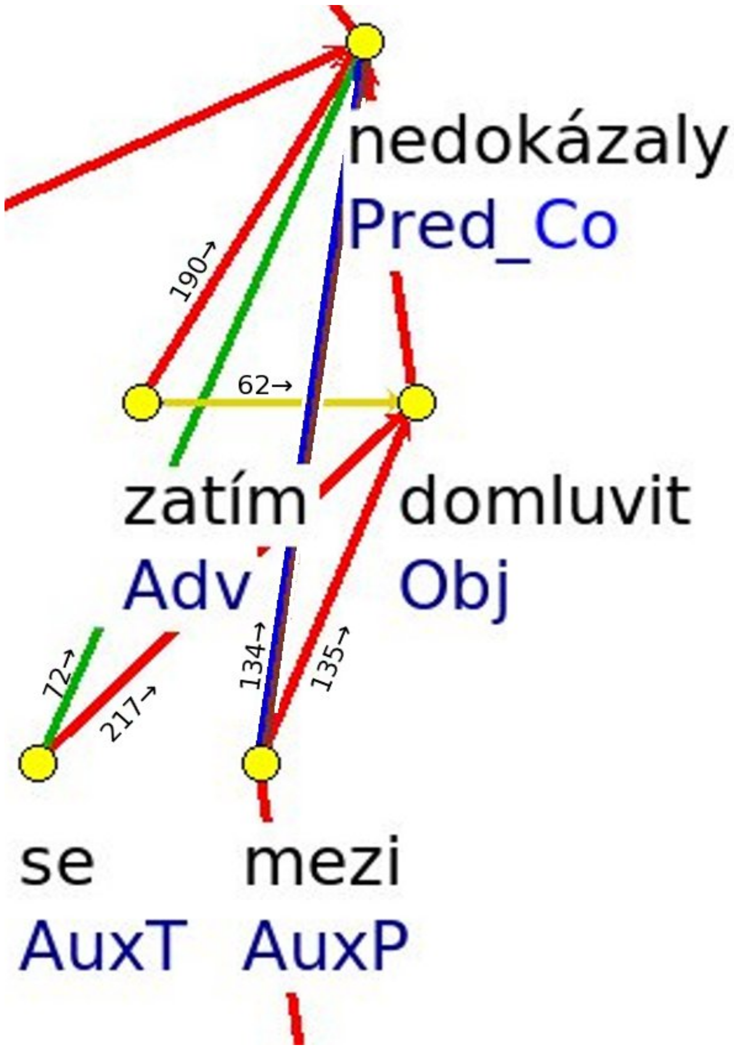
Privatizované mlékárny se však zatím mezi sebou nedokázaly domluvit.

However, the privatized dairies were not yet able to agree among themselves.

			Mate	Turbo	Malt	MST	GOLD
1	Privatizované	privatized	**2**	**2**	**2**	**2**	**2**
2	mlékárny	dairies	**8**	**8**	**8**	**8**	**8**
3	se	*refl.*	**9**	**8**	**9**	**9**	**9**
4	však	however	**0**	**0**	**0**	**0**	**0**
5	zatím	yet	**8**	**8**	**8**	**9**	**8**
6	mezi	among	**8**	**9**	**8**	**9**	**9**
7	sebou	themselves	**6**	**6**	**6**	**6**	**6**
8	nedokázaly	were unable	**4**	**4**	**4**	**4**	**4**
9	domluvit	agree	**8**	**8**	**8**	**8**	**8**
10	.	.	**0**	**0**	**0**	**0**	**0**

Privatizované mlékárny se však zatím mezi sebou nedokázaly domluvit.
However, the privatized dairies were not yet able to agree among themselves.





Agreements of parsers

**Another kind of information extracted:
if parsers A+B predict one dep. relation, parsers X+Y predict another one, parser Z a third one,
how often are A+B correct?**

Mate+Turbo | Malt+MST | Parsito

45.71% (UAS)

Agreements of parsers: example of table entries

Un/lab	Parser(s)	Other parsers	Accur.y	Occurr.
UAS	all agree	none	97.16	867 999
...	...	...	...	...
UAS	Malt+MST+Parsito	Mate+Turbo	47.25	5 543
UAS	Mate+Turbo	Malt+Parsito |MST	46.96	2 329
UAS	Mate+Turbo	Malt+MST | Parsito	45.71	1 426
UAS	Mate+Turbo	Malt+MST+Parsito	44.70	5 543
UAS	Mate+Turbo	Malt | MST | Parsito	44.53	2 237
UAS	Mate+Turbo	Malt | MST+Parsito	44.34	1 449

Results

	UAS	LAS	SENT_U	SENT_L
Mate	88.50	83.03	45.56	33.59
Acc. rates	90.07	84.83	47.61	34.78
Agreements	90.11	80.55	47.67	28.53

Need to revise the algorithm for the choice of dep. label in "agreements"...

What next?

- **test the whole procedure on other languages**
  find best settings for several parsers, optimize
  11 – 110 x train, calculate accuracy rates and/or agreements...
- **create rules for penalizing impossible or very unlikely syntactic structures**
  the weights for (probably) erroneous edges will be penalized, the edges will not be excluded completely
- more varied morphological input (rule based disambiguation, several taggers etc.)

McDonald, R. F. Pereira, K. Ribarov and J. Hajič (2005): Non-projective Dependency Parsing using Spanning Tree Algorithms. In *Proceedings of HLT-EMNLP 2005.*

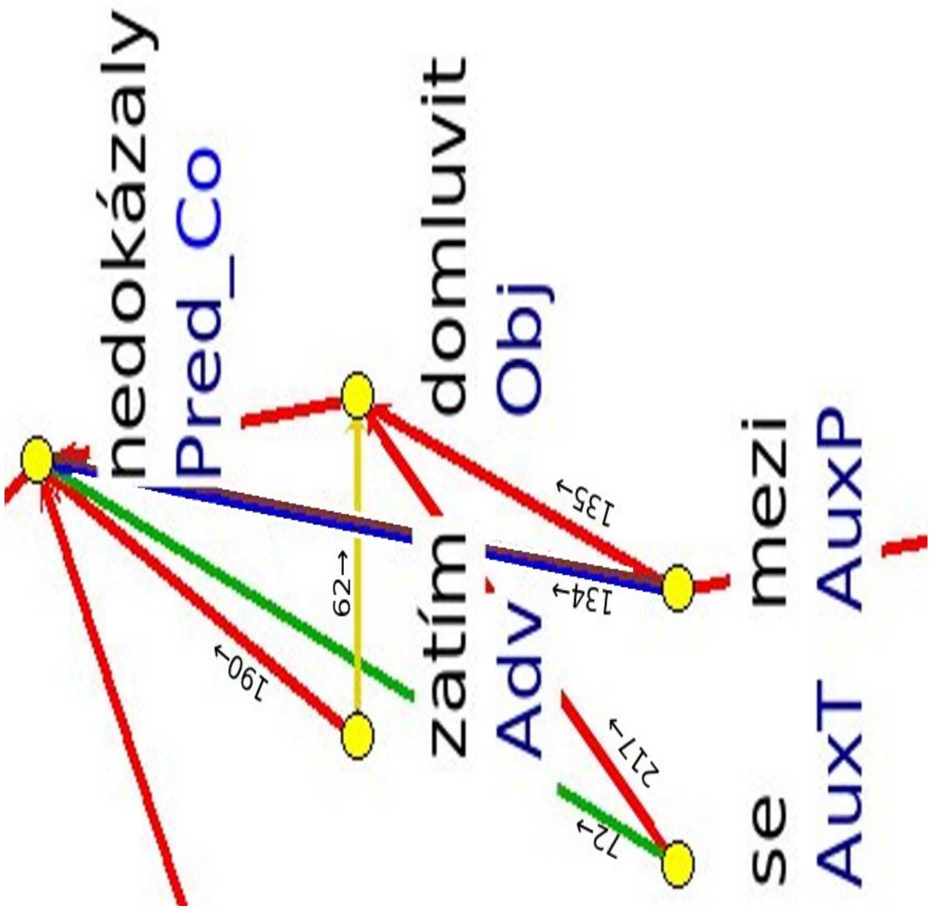
Nivre, J., J. Hall and J. Nilsson (2006): MaltParser: A Data-Driven Parser-Generator for Dependency Parsing. In *Proceedings of LREC 2006*.

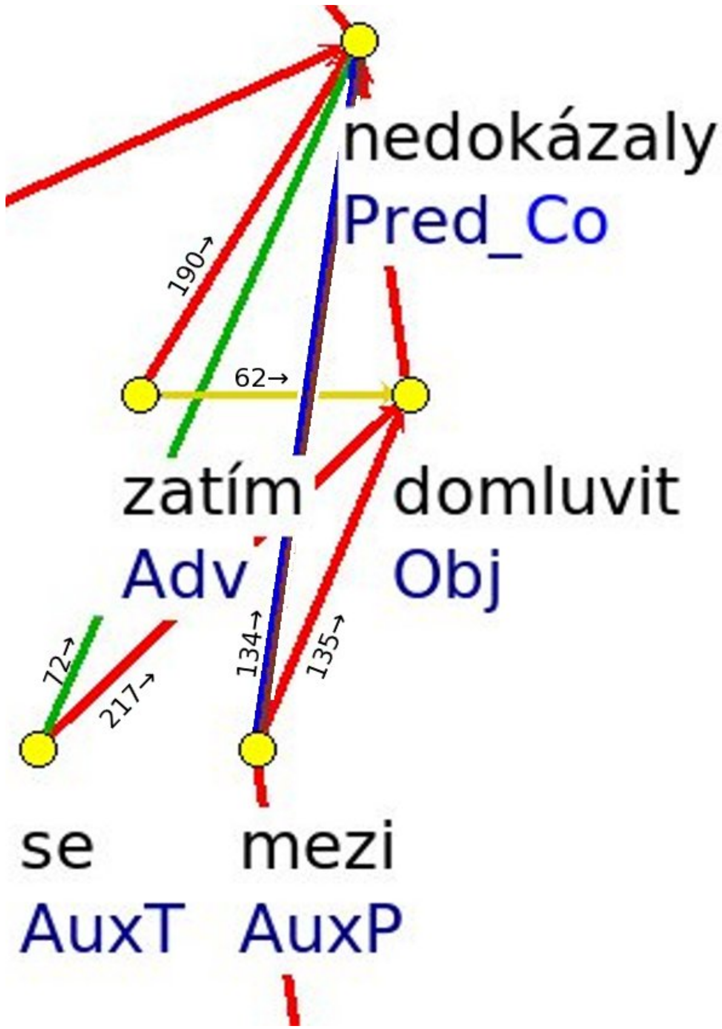
Bohnet, B. and J. Nivre (2012): A Transition-Based System for Joint Part-of-Speech Tagging and Labeled Non-Projective Dependency Parsing. In *Proceedings of EMNLP-CoNLL 2012***.**

Martins, A. F. T., M. B. Almeida and N. A. Smith (2013): Turning on the Turbo: Fast Third-Order Non-Projective Turbo Parsers. In *Proceedings of* ACL *2013*.

Straka M., J. Hajič, J. Straková and J. Hajič jr. (2015): Parsing Universal Dependency Treebanks using Neural Networks and Search-Based Oracle. In *Proceedings of TLT 2015*.

Thank you!


nedokázaly
Pred_Co
domluvit
Obj
zatím
Adv
mezi
AuxP
se
AuxT
190→
62→
135→
134→
217→
72→


nedokázaly
Pred_Co
190→
62→
zatím
Adv
domluvit
Obj
72→
217→
134→
135→
se
AuxT
mezi
AuxP