12 Years of Unsupervised Dependency Parsing

David Mareček

Institute of Formal and Applied Linguistics
Charles University in Prague
Czech Republic

SloNLP 2016, September 18th, Tatranské Matliare, Slovakia

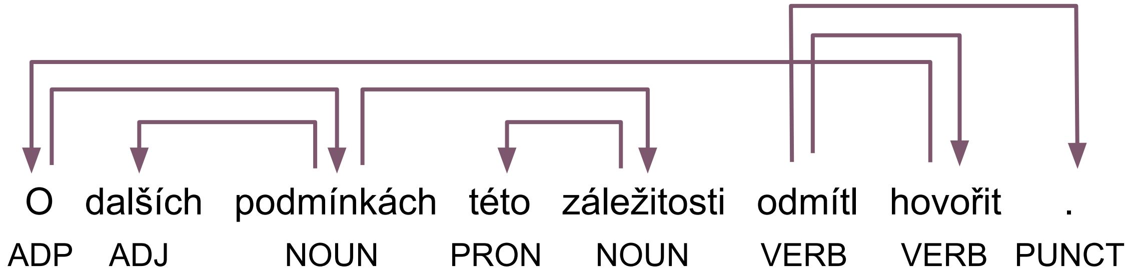
Dependency parsing

O dalších podmínkách této záležitosti odmítl hovořit .

Dependency parsing

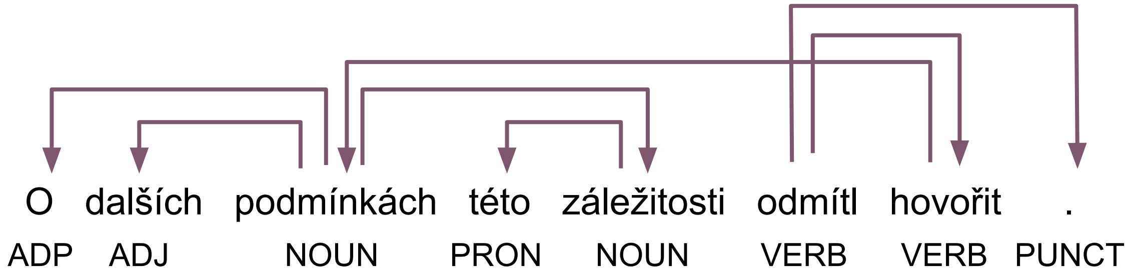
O	dalších	podmínkách	této	záležitosti	odmítl	hovořit	.
ADP	ADJ	NOUN	PRON	NOUN	VERB	VERB	PUNCT

Dependency parsing



PDT style

Dependency parsing



Universal Dependencies style

Outline

- POSSIBILITIES OF PARSING LANGUAGES WITH LIMITED RESOURCES

- From completely supervised to completely unsupervised
  - Completely Supervised
  - Semi-supervised
  - Projection
  - Delexicalization
  - Minimally supervised
  - Unsupervised using POS
  - Completely unsupervised

- Results comparison
- Conclusions

Motivation and Resources

Motivation 1: We want to parse a language, for which no or very small annotated treebanks exists.
Motivation 2: The structures we want to get are different from that we have in the treebank.

Resources for parsing:

- Tagger (manually tagged corpus)
- Dependency treebank
- Only a few annotated sentences
- A parallel corpus with another language, for which we have a treebank
- Shared tagset with another language, for which we have a treebank
- Grammar rules (based on tags)

[en] He declined to discuss other terms of the issue.
[cs] O dalších podmínkách této záležitosti odmítl hovořit.

Different degrees of supervision

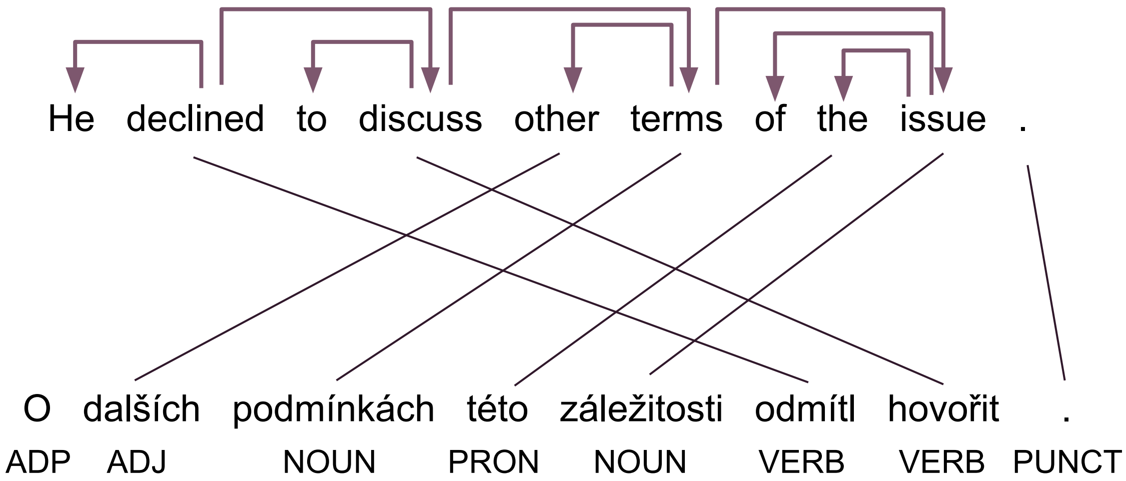
	Tagger for X	Treebank for X	Grammar rules for X	Parallel corpus X - Y	Shared tagset X - Y	Treebank k for Y	Raw text corpus for X
Completely supervised	o	o					
Self-training	o	small					o
Projection method	o			o		o	
Delexicalization method	o				o	o	
Minimally supervised	o		o				o
Unsupervised using POS	o						o
Unsupervised without POS							o

Projection methods

Motivation: We want to parse language (X), for which no treebank exists.

Resources: tagger, parallel treebank (X-Y) with another language (Y), for which we have a parser.

1. Parse the Y side of the parallel treebank X-Y
2. Do the word alignment between X and Y.

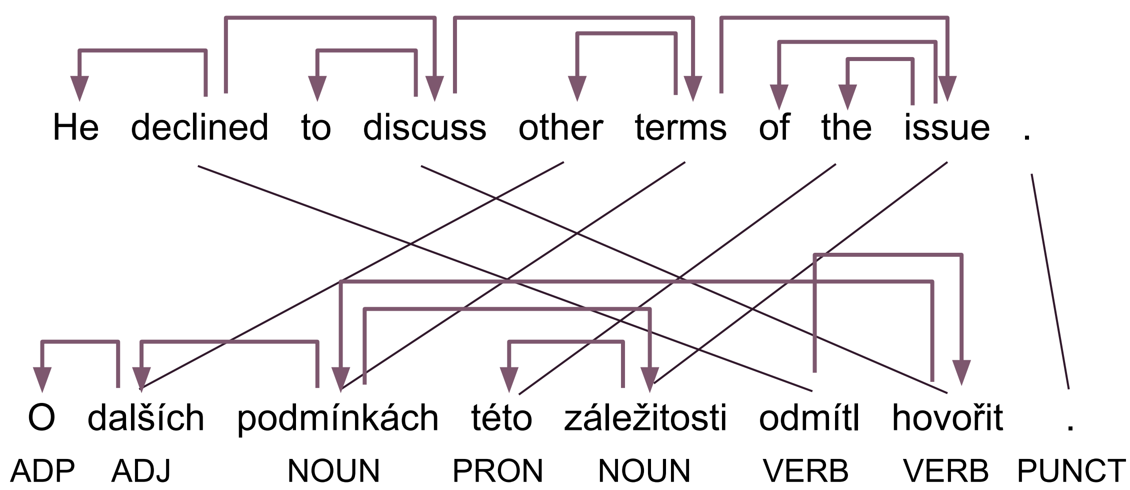


Projection methods

Motivation: We want to parse language (X), for which no treebank exists.

Resources: tagger, parallel treebank (X-Y) with another language (Y), for which we have a parser.

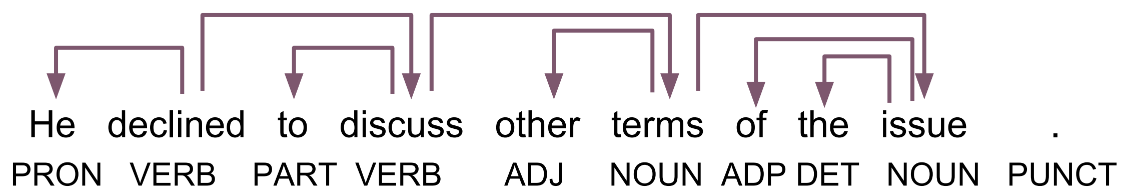
1. Parse the Y side of the parallel treebank X-Y
2. Do the word alignment between X and Y.
3. Project the dependency edges from Y to X.
4. Attach somehow the remaining nodes.
5. Train a parser on the projected trees X.



Delexicalization methods

Motivation: We want to parse language (X), for which no treebank exists.

Resources: tagger, shared tagset with another language (Y), for which we have a treebank.

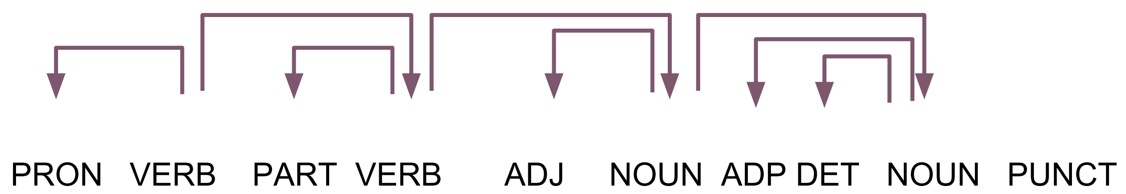


Delexicalization methods

Motivation: We want to parse language (X), for which no treebank exists.

Resources: tagger, shared tagset with another language (Y), for which we have a treebank.

1. Delete wordforms from the treebank Y and train a supervised parser only on its POS tags.

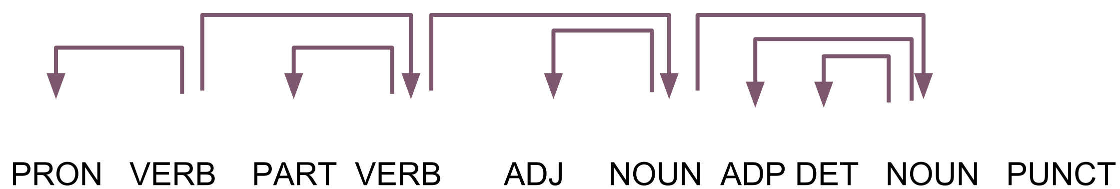


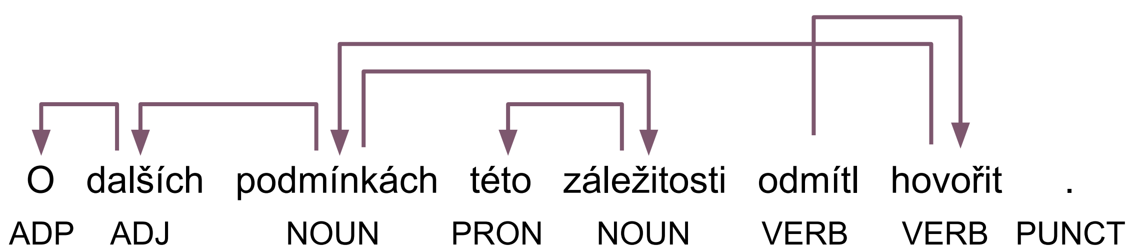
Delexicalization methods

Motivation: We want to parse language (X), for which no treebank exists.

Resources: tagger, shared tagset with another language (Y), for which we have a treebank.

1. Delete wordforms from the treebank Y and train a supervised parser only on its POS tags.
2. Use such parser for language X, use only the POS tags of X. The tagsets must be shared between X and Y.





Unsupervised methods with POS

Motivation: We want to parse language (X), for which no annotated treebank exists. We do not want to imitate any treebank, we want to infer structures only from POS-tagged texts.

- Not burdened by linguistic rules (what to do with coordinations, appositions, complex verbs forms, punctuation, …), everything is learned directly from the corpus
- The structures obtained by unsupervised parsers can be tuned (fitted) to particular applications, while the supervised parsers always simulate the treebanks

Resources: tagger, raw corpus

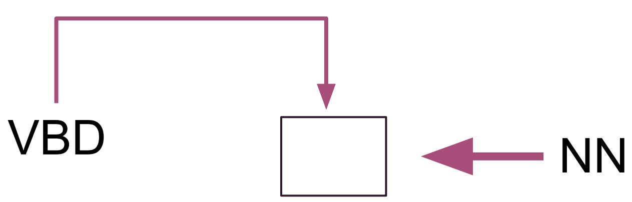
Unsupervised methods with POS

DEPENDENCY MODEL WITH VALENCE

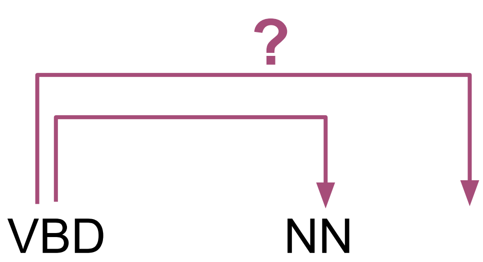
- generative model
    - ***choose*** probability for generating labels of nodes
    - ***stop*** probability for generating dependency edges
- introduced by Klein and Manning (2004)
- improved by Smith (2007), Headden (2009), Spitkovsky (2010-2012), ...

$P_{choose}(c_d|c_h, dir)$

$P_{stop}(STOP|c_h, dir, adj)$



P_{choose}(NN | VBD, right)



$P_{stop}(\neg$ STOP | VBD, right, 0)

Unsupervised methods with POS

BAYESIAN INFERENCE - GIBBS SAMPLING

1. *initialization* - random projective trees

2. *sampling* - In many iterations, we choose one sentence and

 - compute probability of each possible projective dependency tree using dynamic programming

 - sample a new tree according to the computed probability distribution

3. *finalization* - an averaged trees during the sampling are outputed

$$P_{choose}(c_d|c_h, dir) = \frac{\frac{1}{|C|}\alpha_c + count^{-}(c_d, c_h, dir)}{\alpha_c + count^{-}(c_h, dir)}$$

$$P_{stop}(STOP|c_h, dir, adj) = \frac{\frac{2}{3}\alpha_s + count^{-}(STOP, c_h, dir, adj)}{\alpha_s + count^{-}(c_h, dir, adj)}$$

Minimally supervised methods

We add a couple of linguistic rules to guide the unsupervised parsing, e.g:

- function words (tags ADP, DET, AUX, CONJ, SCONJ, PUNCT, ...) have no children
- ADJ dependens often on NOUNs
- VERBs are in roots.

The structures are then much closer to the gold dependency trees.

Unsupervised methods without POS

If we do not have any tagger:

We can run a word-clustering tool to induce a class for each word.

And then run an unsupervised parser on these classes instead of POS tags.

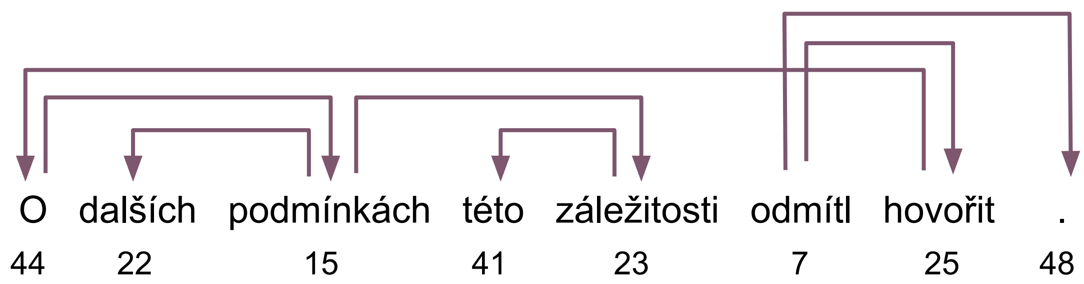
O	dalších	podmínkách	této	záležitosti	odmítl	hovořit	.
44	22	15	41	23	7	25	48

Unsupervised methods without POS

If we do not have any tagger:

We can run a word-clustering tool to induce a class for each word.

And then run an unsupervised parser on these classes instead of POS tags.



Different degrees of supervision

	Tagger for X	Treebank for X	Grammar rules for X	Parallel corpus X - Y	Shared tagset X - Y	Treebank for Y	Raw text corpus for X
Completely supervised	o	o					
Self-training	o	small					o
Projection method	o			o		o	
Delexicalization method	o				o	o	
Minimally supervised	o		o				o
Unsupervised using POS	o						o
Unsupervised without POS							o

Results

Unlabelled attachment score on selected languages from CoNLL 2006 and 2007 datasets

	bg	cs	de	el	en	es	hu	it	pt	sv	AVG
Completely supervised	87.4	86.3	87.3	84.0	90.1	82.3	83.6	87.9	87.6	84.6	86.1
Projection method	56.3	62.3	50.1	65.2	--	58.2	51.2	59.3	62.9	55.8	57.9
Delexicalization method	60.3	57.5	51.7	58.5	62.4	55.6	58.0	56.8	67.7	58.3	58.7
Minimally supervised	58.1	54.8	53.9	53.2	56.6	62.7	57.4	54.8	69.8	59.5	56.9
Unsupervised using POS	54.9	52.4	52.4	26.3	55.4	61.1	34.0	39.4	69.6	54.5	50.0
Unsupervised without POS	47.9	38.0	41.2	39.7	47.9	60.1	24.1	41.4	31.4	54.9	42.6

Conclusions - 1

Unlabelled attachment score on selected languages from CoNLL 2006 and 2007 datasets

	bg	cs	de	el	en	es	hu	it	pt	sv	AVG
Completely supervised	87.4	86.3	87.3	84.0	90.1	82.3	83.6	87.9	87.6	84.6	86.1
Projection method	56.3	62.3	50.1	65.2	--	58.2	51.2	59.3	62.9	55.8	57.9
Delexicalization method	60.3	57.5	51.7	58.5	62.4	55.6	58.0	56.8	67.7	58.3	58.7
Minimally supervised	58.1	54.8	53.9	53.2	56.6	62.7	57.4	54.8	69.8	59.5	56.9
Unsupervised using POS	54.9	52.4	52.4	26.3	55.4	61.1	34.0	39.4	69.6	54.5	50.0
Unsupervised without POS	47.9	38.0	41.2	39.7	47.9	60.1	24.1	41.4	31.4	54.9	42.6

Minimally supervised parsing is always better than unsupervised
[Mareček @ SLSP 2016]

Conclusions - 2

Unlabelled attachment score on selected languages from CoNLL 2006 and 2007 datasets

	bg	cs	de	el	en	es	hu	it	pt	sv	AVG
Completely supervised	87.4	86.3	87.3	84.0	90.1	82.3	83.6	87.9	87.6	84.6	86.1
Projection method	56.3	62.3	50.1	65.2	--	58.2	51.2	59.3	62.9	55.8	57.9
Delexicalization method	60.3	57.5	51.7	58.5	62.4	55.6	58.0	56.8	67.7	58.3	58.7
Minimally supervised	58.1	54.8	53.9	53.2	56.6	62.7	57.4	54.8	69.8	59.5	56.9
Unsupervised using POS	54.9	52.4	52.4	26.3	55.4	61.1	34.0	39.4	69.6	54.5	50.0
Unsupervised without POS	47.9	38.0	41.2	39.7	47.9	60.1	24.1	41.4	31.4	54.9	42.6

Delexicalized parsing is a simple method with reasonable results, if you want to transfer the syntax style from another language.

Conclusions - 3

Unlabelled attachment score on selected languages from CoNLL 2006 and 2007 datasets

	bg	cs	de	el	en	es	hu	it	pt	sv	AVG
Completely supervised	87.4	86.3	87.3	84.0	90.1	82.3	83.6	87.9	87.6	84.6	86.1
Projection method	56.3	62.3	50.1	65.2	--	58.2	51.2	59.3	62.9	55.8	57.9
Delexicalization method	60.3	57.5	51.7	58.5	62.4	55.6	58.0	56.8	67.7	58.3	58.7
Minimally supervised	58.1	54.8	53.9	53.2	56.6	62.7	57.4	54.8	69.8	59.5	56.9
Unsupervised using POS	54.9	52.4	52.4	26.3	55.4	61.1	34.0	39.4	69.6	54.5	50.0
Unsupervised without POS	47.9	38.0	41.2	39.7	47.9	60.1	24.1	41.4	31.4	54.9	42.6

Fully unsupervised parsing is very interesting problem, however currently without any obvious application.

Thank you for your attention!

http://ufal.mff.cuni.cz/udp/