Universal Dependencies are hard to parse – or are they?

Ines Rehbein, Julius Steen, Bich-Ngoc Do & Anette Frank

DepLing 2017

Empirical Linguistics
and Computational Language Modeling
Leibniz ScienceCampus

Motivation

Universal Dependencies (UD) provide the means for

- a cross-linguistically consistent dependency annotation
- good for multilingual applications

Motivation

Universal Dependencies (UD) provide the means for

- a cross-linguistically consistent dependency annotation
- good for multilingual applications

But ...

- not optimised for monolingual dependency parsing
  - parsers prefer function-head encodings / chain-like representations

  $\rightarrow$ lower parsing accuracies for UD-style encoding

- What causes the decrease in parsing accuracy?
- Is it equally strong for typologically different languages?

Related work

- Schwartz et al. (2012), Popel et al. (2013), Maraček et al (2013), Versley & Kirilin (2015), Rosa (2015), de Lhoneux & Nivre (2016), ...

Related work

- Schwartz et al. (2012), Popel et al. (2013), Maraček et al (2013), Versley & Kirilin (2015), Rosa (2015), de Lhoneux & Nivre (2016), ...

- Silveira and Manning (2015)
  - convert English content-head to function-head representations
  - $\rightarrow$ improved acc. for function-head representation of PPs and copula
  - back-conversion of parser output propagates errors

Related work

- Schwartz et al. (2012), Popel et al. (2013), Maraček et al (2013), Versley & Kirilin (2015), Rosa (2015), de Lhoneux & Nivre (2016), ...

- Silveira and Manning (2015)
  - convert English content-head to function-head representations
  - → improved acc. for function-head representation of PPs and copula
  - back-conversion of parser output propagates errors

- Kohita et al. (2017)
  - convert 3 design decisions in UD (*case, dep, mark*) to function-head
  - conversion improves results of *back-converted* parser output for 16 out of 19 languages

Open questions

- Impact of parsing model/feature templates
  - do different parsing paradigms favour specific annotation schemes?
  - are feature templates biased towards one particular representation?
- Typological properties
  - same trend for typologically different languages?

Open questions

- Impact of parsing model/feature templates
  - do different parsing paradigms favour specific annotation schemes?
  - are feature templates biased towards one particular representation?
- Typological properties
  - same trend for typologically different languages?

Method:

- Train and evaluate different parsing models on original UD and on converted treebanks for 15 languages
- Conversion of preposition, copula and coordination handling

Data

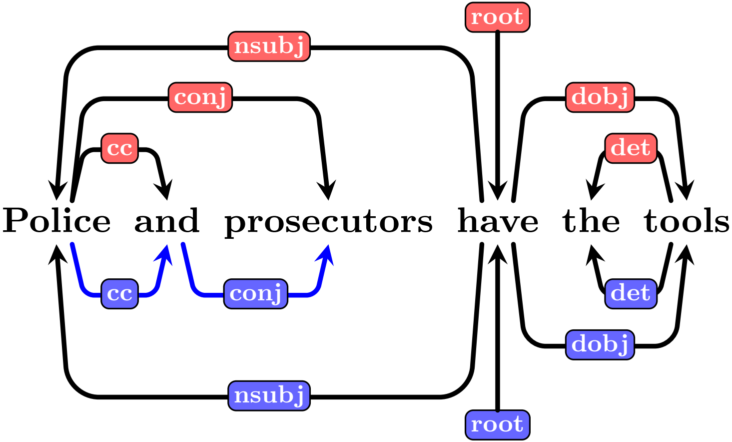
- 15 languages from the UD project (v1.3)
  - treebank sizes between 3,900 and 68,500 trees
  - treebanks that have undergone some manual quality check

germanic	*iranian*	*romance*	*slavic*	*sinitic*	*finnic*	*turkic*
English	Farsi	Catalan	Bulgarian	Chinese	Estonian	Turkish
German		Spanish	Czech			
		French	Hungarian			
		Italian	Russian			
		Romanian				

Conversions

Coordination

UD encoding

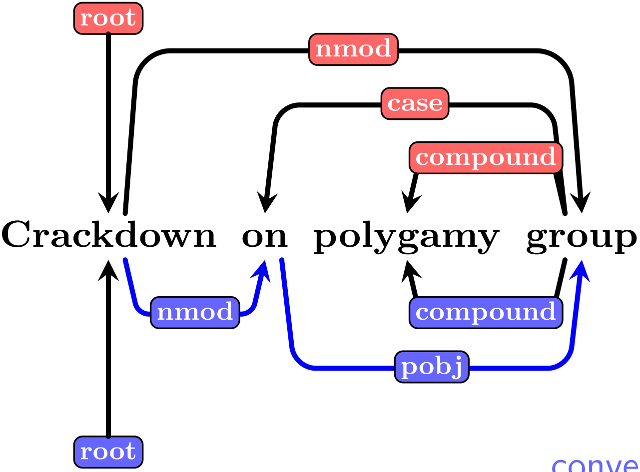


converted tree

Conversions

Preposition

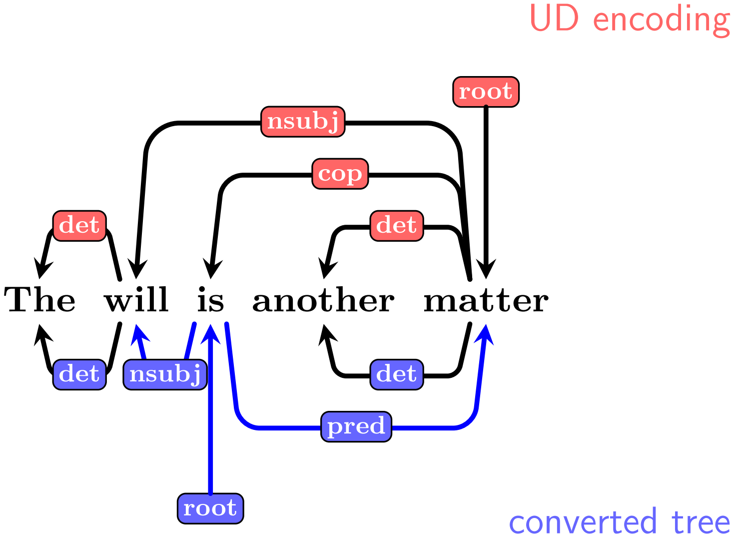
UD encoding



converted tree

Conversions

Copula



Conversions

Evaluation

- Original UD treebanks
  - train parser on gold UD trees
  - evaluate UD parser output against original UD trees
- Converted treebanks
  - train parser on converted UD trees
  - evaluate parser output against converted gold trees

  → avoid errors introduced by back-conversion of parser output

Information preservation after back-and-forth conversion on gold trees

		% affected	size	LAS cop	LAS prep	LAS coord	LAS c-p-c	UAS c-p-c
Chinese	**zh**	20.9	3,997	100.0	100.0	99.9	99.9	100.0
Estonian	**et**	23.6	14,510	99.9	100.0	100.0	99.9	100.0
Turkish	**tr**	27.9	3,948	99.9	99.8	99.8	99.4	99.8
Russian	**ru**	30.6	48,171	100.0	100.0	100.0	100.0	100.0
German	**de**	33.2	14,118	99.8	100.0	99.8	99.6	100.0
Czech	**cs**	35.3	68,495	100.0	100.0	99.7	99.7	100.0
Romanian	**ro**	36.4	7,141	99.9	99.9	99.8	99.7	100.0
English	**en**	37.6	12,543	100.0	99.8	99.9	99.6	99.9
Croatian	**hr**	38.5	5,792	100.0	100.0	99.8	99.8	100.0
French	**fr**	38.5	14,554	100.0	99.8	99.9	99.8	99.9
Catalan	**ca**	38.8	13,123	99.9	99.5	99.9	99.4	99.8
Italian	**it**	40.3	12,837	100.0	100.0	99.9	100.0	100.0
Spanish	**es**	40.3	14,187	99.8	99.9	99.9	99.6	99.9
Bulgarian	**bg**	43.7	8,907	100.0	100.0	99.9	99.9	100.0
Farsi	**fa**	45.7	4,798	99.6	100.0	98.8	98.4	100.0
avg.		*35.4*	*16,475*	*99.9*	*99.9*	*99.8*	*99.6*	*99.9*

Information preservation after back-and-forth conversion on gold trees

		% affected	size	LAS				UAS
				cop	prep	coord	c-p-c	c-p-c
Chinese	**zh**	20.9	3,997	100.0	100.0	99.9	99.9	100.0
Estonian	**et**	23.6	14,510	99.9	100.0	100.0	99.9	100.0
Turkish	**tr**	27.9	3,948	99.9	99.8	99.8	99.4	99.8
Russian	**ru**	30.6	48,171	100.0	100.0	100.0	100.0	100.0
German	**de**	33.2	14,118	99.8	100.0	99.8	99.6	100.0
Czech	**cs**	35.3	68,495	100.0	100.0	99.7	99.7	100.0
Romanian	**ro**	36.4	7,141	99.9	99.9	99.8	99.7	100.0
English	**en**	37.6	12,543	100.0	99.8	99.9	99.6	99.9
Croatian	**hr**	38.5	5,792	100.0	100.0	99.8	99.8	100.0
French	**fr**	38.5	14,554	100.0	99.8	99.9	99.8	99.9
Catalan	**ca**	38.8	13,123	99.9	99.5	99.9	99.4	99.8
Italian	**it**	40.3	12,837	100.0	100.0	99.9	100.0	100.0
Spanish	**es**	40.3	14,187	99.8	99.9	99.9	99.6	99.9
Bulgarian	**bg**	43.7	8,907	100.0	100.0	99.9	99.9	100.0
Farsi	**fa**	45.7	4,798	99.6	100.0	98.8	98.4	100.0
avg.		*35.4*	*16,475*	*99.9*	*99.9*	*99.8*	*99.6*	*99.9*

Kohita et al. 2017: 2.3 - 15.6%

Information preservation after back-and-forth conversion on gold trees

			size	LAS				UAS
		% affected		cop	prep	coord	c-p-c	c-p-c
Chinese	**zh**	20.9	3,997	100.0	100.0	99.9	99.9	100.0
Estonian	**et**	23.6	14,510	99.9	100.0	100.0	99.9	100.0
Turkish	**tr**	27.9	3,948	99.9	99.8	99.8	99.4	99.8
Russian	**ru**	30.6	48,171	100.0	100.0	100.0	100.0	100.0
German	**de**	33.2	14,118	99.8	100.0	99.8	99.6	100.0
Czech	**cs**	35.3	68,495	100.0	100.0	99.7	99.7	100.0
Romanian	**ro**	36.4	7,141	99.9	99.9	99.8	99.7	100.0
English	**en**	37.6	12,543	100.0	99.8	99.9	99.6	99.9
Croatian	**hr**	38.5	5,792	100.0	100.0	99.8	99.8	100.0
French	**fr**	38.5	14,554	100.0	99.8	99.9	99.8	99.9
Catalan	**ca**	38.8	13,123	99.9	99.5	99.9	99.4	99.8
Italian	**it**	40.3	12,837	100.0	100.0	99.9	100.0	100.0
Spanish	**es**	40.3	14,187	99.8	99.9	99.9	99.6	99.9
Bulgarian	**bg**	43.7	8,907	100.0	100.0	99.9	99.9	100.0
Farsi	**fa**	45.7	4,798	99.6	100.0	98.8	98.4	100.0
avg.		*35.4*	*16,475*	*99.9*	*99.9*	*99.8*	*99.6*	*99.9*

Experimental setup

- Three different non-projective parsers

RBG parser	[Lei et al. 2014]	*graph-based, 3rd-order features*
IMSTrans	[Björkelund & Nivre 2015]	*transition-based, standard features*
DENSE parser	[Zhang et al. 2017]	*bi-LSTM-based, no feature templates*

- Evaluation measures
  - Labelled Attachment Score (LAS)
  - Content dependencies only (CNC) [Nivre & Fang 2017]

Baselines

Parsing scores for UD encodings (no conversion)

		CNC		
		IMS	RBG	DENSE
germanic	de	**79.7**	78.9	77.1
	en	**82.8**	82.2	82.3
iranian	fa	80.5	79.5	**80.8**
romance	ca	**84.0**	82.7	83.6
	es	**78.6**	77.5	78.0
	fr	**79.4**	77.6	78.6
	it	**84.3**	82.9	83.9
	ro	**75.4**	74.6	73.3
slavic	bg	**83.7**	80.8	81.7
	cs	**86.1**	83.9	83.5
	hr	77.2	**77.6**	74.9
	ru	**88.0**	87.8	84.4
sinitic	zh	**80.6**	77.9	79.1
finnic	et	**83.0**	82.6	73.0
turkic	tr	71.9	**73.4**	59.1

Results for converted treebanks

		IMS		RBG		DENSE	
	lang	CNC	Δ	CNC	Δ	CNC	Δ
ger	de	81.0		**81.2**		78.0	
	en	**83.6**		83.4		**83.6**	
ira	fa	**84.2**		83.4		83.6	
rom	ca	**85.6**		85.0		84.9	
	es	80.5		**80.8**		79.9	
	fr	81.9		80.7		80.4	
	it	**86.1**		**86.1**		85.5	
	ro	**75.7**		75.3		73.6	
sla	bg	**85.4**		83.8		83.8	
	cs	**87.3**		85.2		84.2	
	hr	**77.4**		77.3		73.2	
	ru	**89.2**		88.7		82.1	
sin	zh	**81.9**		78.9		79.2	
fin	et	**84.4**		82.8		74.7	
tur	tr	71.6		**71.8**		58.3	

CNC and differences (Δ) to CNC obtained on original UD treebanks

Results for converted treebanks

		IMS		RBG		DENSE	
	lang	CNC	Δ	CNC	Δ	CNC	Δ
ger	de	81.0	**1.3**	81.2	**2.3**	78.0	0.9
	en	83.6	0.8	83.4	**1.2**	83.6	**1.3**
ira	fa	84.2	**3.7**	83.4	**3.9**	83.6	**2.8**
rom	ca	85.6	**1.6**	85.0	**2.3**	84.9	**1.3**
	es	80.5	**1.9**	80.8	**3.3**	79.9	**1.9**
	fr	81.9	**2.5**	80.7	**3.1**	80.4	**1.8**
	it	86.1	**1.8**	86.1	**3.2**	85.5	**1.6**
	ro	75.7	0.3	75.3	0.7	73.6	0.3
sla	bg	85.4	**1.7**	83.8	**3.0**	83.8	**2.1**
	cs	87.3	**1.2**	85.2	**1.3**	84.2	0.7
	hr	77.4	0.2	77.3	**-0.3**	73.2	**-1.7**
	ru	89.2	**1.2**	88.7	0.9	82.1	**-2.3**
sin	zh	81.9	**1.3**	78.9	**1.0**	79.2	0.1
fin	et	84.4	**1.4**	82.8	0.2	74.7	**1.7**
tur	tr	71.6	**-0.3**	71.8	**-1.6**	58.3	**-0.8**

CNC and differences (Δ) to CNC obtained on original UD treebanks

Results for converted treebanks

	lang	IMS CNC	IMS Δ	RBG CNC	RBG Δ	DENSE CNC	DENSE Δ
ger	de	81.0	**1.3**	81.2	**2.3**	78.0	0.9
	en	83.6	0.8	83.4	**1.2**	83.6	**1.3**
ira	fa	84.2	**3.7**	83.4	**3.9**	83.6	**2.8**
rom	ca	85.6	**1.6**	85.0	**2.3**	84.9	**1.3**
	es	80.5	**1.9**	80.8	**3.3**	79.9	**1.9**
	fr	81.9	**2.5**	80.7	**3.1**	80.4	**1.8**
	it	86.1	**1.8**	86.1	**3.2**	85.5	**1.6**
	ro	75.7	0.3	75.3	0.7	73.6	0.3
sla	bg	85.4	**1.7**	83.8	**3.0**	83.8	**2.1**
	cs	87.3	**1.2**	85.2	**1.3**	84.2	0.7
	hr	77.4	0.2	77.3	**-0.3**	73.2	**-1.7**
	ru	89.2	**1.2**	88.7	0.9	82.1	**-2.3**
sin	zh	81.9	**1.3**	78.9	**1.0**	79.2	0.1
fin	et	84.4	**1.4**	82.8	0.2	74.7	**1.7**
tur	tr	71.6	**-0.3**	71.8	**-1.6**	58.3	**-0.8**

No feature template bias!

What determines ...

whether and how much a specific language will benefit from a particular choice of encoding?

- amount of non-projectivity
- morphological richness
- ambiguity caused by case syncretism
- word order properties

What determines ...

whether and how much a specific language will benefit from a particular choice of encoding?

- amount of non-projectivity
- morphological richness
- ambiguity caused by case syncretism
- word order properties
  - dependency length
  - ambiguity in head direction

[Gulordava & Merlo 2016]

Impact on parsing accuracies using *artificial* treebanks

Gulordava & Merlo 2016

- We adopt their measures to find out more about the **impact of annotation schemes** on parsing
  - Compute overall ratio of dependency length minimisation (DLM) in the converted treebanks *(variation in dependency length)*
  - Compute arc direction entropy (ADE) *(variation in linear order)*

Dependency length minimisation (DLM ratio)

$$\mathit{DLM\ ratio} = \sum_s \frac{DL_s}{|s|^2} / \sum_s \frac{ModDL_s}{|s|^2} \tag{1}$$

- *DL*: sum of length of all arcs in the tree for sentence *s*
- DLM ratio > 1: conversion resulted in a decrease in DL

DLM ratio for individual conversions

	Lang	DL (orig)	DLM (conv)
ger	de	3.4	1.03
	en	2.9	**1.07**
ira	fa	3.5	0.97
rom	ca	3.1	**1.09**
	es	2.8	**1.11**
	fr	2.8	**1.09**
	it	2.7	**1.08**
	ro	2.7	**1.07**
sla	bg	2.5	**1.08**
	cs	2.8	**1.06**
	hr	2.8	**1.08**
	ru	2.7	**1.05**
sin	zh	3.6	1.00
fin	et	2.6	1.02
tur	tr	2.6	1.02

Table: Avg. DL in the original treebank and DLM ratio for converted trees

Impact of dependency length minimisation on parsing

- For most languages conversion reduces dependency length
  - no effect for Chinese
  - longer dependencies for Farsi

Impact of dependency length minimisation on parsing

- For most languages conversion reduces dependency length
  - no effect for Chinese
  - longer dependencies for Farsi

Linear Regression

- fit linear regression model to the data

$\rightarrow$ no significant correlation between DL and changes in CNC

(IMSTrans: $p = 0.604$, RBG: $p = 0.463$, DENSE: $p = 0.943$)

Impact of dependency length minimisation on parsing

- For most languages conversion reduces dependency length
  - no effect for Chinese
  - longer dependencies for Farsi

Linear Regression

- fit linear regression model to the data

→ no significant correlation between DL and changes in CNC

(IMSTrans: $p = 0.604$, RBG: $p = 0.463$, DENSE: $p = 0.943$)

Improvements in parsing accuracies not due to shorter dependencies!

Arc direction entropy (ADE)

- Impact of variation in linear ordering of head and dependent on parsing accuracy?
  - Typological differences between languages:
    head-initial vs. head-final (Liu 2010)
  - Annotation schemes also impact head direction

Arc direction entropy (ADE)

- Impact of variation in linear ordering of head and dependent on parsing accuracy?
  - Typological differences between languages:
    head-initial vs. head-final (Liu 2010)
  - Annotation schemes also impact head direction

- We compute ADE, following Gulordava & Merlo 2016:
  calculate overall arc direction entropy in a treebank conditioned on relation type *Rel* and POS for head *H* and dependent *D*

$$H(Dir|Rel,H,D) = \sum_{rel,h,d} p(rel,h,d) H(Dir|rel,h,d) \qquad (2)$$

Arc direction entropy (ADE)

- Impact of variation in linear ordering of head and dependent on parsing accuracy?
  - Typological differences between languages: head-initial vs. head-final (Liu 2010)
  - Annotation schemes also impact head direction

- We compute ADE, following Gulordava & Merlo 2016: calculate overall arc direction entropy in a treebank conditioned on relation type *Rel* and POS for head *H* and dependent *D*

$$H(\mathit{Dir}|\mathit{Rel}, H, D) = \sum\nolimits_{rel,h,d} p(rel,h,d) H(Dir|rel,h,d) \tag{2}$$

- Increase in ADE: converted treebank introduced more variation with respect to linear order of head and dependent

Arc direction entropy (ADE)

Methodological issues

Entropy is sensitive to sample size (Futrell et al. 2015)

We measure variation in arc direction over n equally-sized random samples from each treebank (with replacement, $n = 1000$), and report the average over all samples.

ADE for converted treebanks

	lang	Δ cop-prep-coord
ger	de	-0.23
	en	-0.72
ira	fa	-0.60
rom	ca	0.16
	es	-0.36
	fr	-0.27
	it	-0.29
	ro	0.09
sla	bg	-0.34
	cs	0.03
	hr	0.41
	ru	0.41
sin	zh	-0.19
fin	et	-0.16
tur	tr	0.50

Difference (Δ) between avg. unlexicalised arc direction entropy (ADE) in the original treebank and in the modified treebanks

Impact of arc direction entropy on parsing

- Treebank conversion:
  - for most languages conversion decreases ADE
  - slight increase for Czech, Romanian, Catalan
  - substantial increase for Croatian, Russian, Turkish

Impact of arc direction entropy on parsing

- Treebank conversion:
  - for most languages conversion decreases ADE
  - slight increase for Czech, Romanian, Catalan
  - substantial increase for Croatian, Russian, Turkish

Linear Regression

$\rightarrow$ significant correlation between ADE and changes in CNC

(IMSTrans: $p = 0.01$, RBG: $p = 0.04$, DENSE: $p = 0.0002$)

Summary & Discussion

- UD-style encoding results in longer dependencies and higher arc direction entropy for many languages
- Our finding suggest that not the increase in dependency length but the *increase in arc direction entropy* causes lower scores for UD

Summary & Discussion

- UD-style encoding results in longer dependencies and higher arc direction entropy for many languages
- Our finding suggest that not the increase in dependency length but the *increase in arc direction entropy* causes lower scores for UD

- Kohita et al. (2017) discuss *head word vocabulary entropy* as potential reason for higher error rate for UD annotations
- No significant correlation between parsing accuracy (RBG) and head word vocabulary entropy

Summary & Discussion

- UD-style encoding results in longer dependencies and higher arc direction entropy for many languages
- Our finding suggest that not the increase in dependency length but the *increase in arc direction entropy* causes lower scores for UD

- Kohita et al. (2017) discuss *head word vocabulary entropy* as potential reason for higher error rate for UD annotations
- No significant correlation between parsing accuracy (RBG) and head word vocabulary entropy

- Results suggest an interaction between typological properties and effect strength of improvements obtained through treebank conversion

Thanks for listening!

References

- Anders Björkelund and Joakim Nivre. 2015. Non-deterministic oracles for unrestricted non-projective transition-based dependency parsing. In Proceedings of IWPT'15.
- Miryam de Lhoneux and Joakim Nivre. 2016. Should have, would have, could have. Investigating verb group representations for parsing with universal dependencies. In Proceedings of the Workshop on Multilingual and Cross-lingual Methods in NLP.
- Kristina Gulordava and Paola Merlo. 2016. Multilingual dependency parsing evaluation: a large-scale analysis of word order properties using artificial data. Transactions of the Association for Computational Linguistics, 4:343–356.
- Tao Lei, Yu Xin, Yuan Zhang, Regina Barzilay, and Tommi Jaakkola. 2014. Low-rank tensors for scoring dependency structures. In Proceedings of ACL'14.
- Ryosuke Kohita, Hiroshi Noji, and Yuji Matsumoto. 2017. Multilingual back-and-forth conversion between content and function head for easy dependency parsing. In Proceedings of EACL'17.
- David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Daniel Zeman, Zdeněk Žabokrtský, and Jan Hajič. 2013. Cross-language study on influence of coordination style on dependency parsing performance. Technical report, UFAL.
- Joakim Nivre and Chiao-Ting Fang. 2017. Universal dependency evaluation. In Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies.
- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. In Proceedings of LREC'16.
- Joakim Nivre. 2016. Universal dependency evaluation. Technical report, Uppsala University, Sweden.
- Martin Popel, David Mareček , Jan Štěpánek, Daniel Zeman, and Zdenek Zabokrtský. 2013. Coordination structures in dependency treebanks. In Proceedings of EACL'13.
- Rudolf Rosa. 2015. Multi-source cross-lingual delexicalized parser transfer: Prague or Stanford? In Proceedings of DepLing'15.
- Roy Schwartz, Omri Abend, and Ari Rappoport. 2012. Learnability-based syntactic annotation design. In Proceedings of COLING'12.
- Natalia Silveira and Christopher D. Manning. 2015. Does universal dependencies need a parsing representation? an investigation of english. In Proceedings of DepLing'15.
- Yannick Versley and Angelika Kirilin. 2015. What is hard in universal dependency parsing? In Proceedings of SPMRL '15.
- Xingxing Zhang, Jianpeng Cheng, and Mirella Lapata. 2017. Dependency parsing as head selection. In Proceedings of EACL'17.

Backup slides

Baselines

Parsing scores for UD encodings (no conversion)

		LAS			CNC		
		IMS	RBG	DEN	IMS	RBG	DEN
germanic	de	**84.3**	83.8	82.0	**79.7**	78.9	77.1
	en	**86.4**	86.3	86.0	**82.8**	82.2	82.3
iranian	fa	83.4	83.1	**83.9**	80.5	79.5	**80.8**
romance	ca	**89.5**	88.8	89.1	**84.0**	82.7	83.6
	es	**85.6**	85.2	85.2	**78.6**	77.5	78.0
	fr	**85.6**	84.4	85.2	**79.4**	77.6	78.6
	it	**89.6**	88.8	89.3	**84.3**	82.9	83.9
	ro	**79.9**	79.6	78.6	**75.4**	74.6	73.3
slavic	bg	**86.9**	84.9	85.6	**83.7**	80.8	81.7
	cs	**87.8**	86.1	85.7	**86.1**	83.9	83.5
	hr	79.9	**80.7**	78.1	77.2	**77.6**	74.9
	ru	**89.5**	**89.5**	86.8	**88.0**	87.8	84.4
sinitic	zh	**81.8**	79.4	80.4	**80.6**	77.9	79.1
finnic	et	**84.1**	83.9	75.3	**83.0**	82.6	73.0
turkic	tr	73.5	**75.1**	62.5	71.9	**73.4**	59.1

Turkish and Croatian: a closer look

metric	orig	cop	prep	coord	c-p-c	Δ
			Turkish			
CNC	73.4	72.9	72.6	71.9	71.8	**-1.6**
core	65.9	65.3	65.9	64.7	**67.1**	+1.2
non-core	75.5	74.9	74.4	73.9	73.2	-2.3
func	85.6	84.2	83.4	**88.2**	**86.0**	+0.4
			Croatian			
CNC	77.7	75.5	76.9	**78.6**	77.3	**-0.4**
core	81.1	**81.5**	81.0	**81.7**	**81.9**	+0.7
non-core	76.8	74.0	75.9	**77.8**	76.1	-0.9
func	88.5	87.9	87.9	**89.1**	**88.7**	+0.2

Results for different label sets for Turkish and Croatian (RBG parser) and difference (Δ) between original and converted treebank (**c**op-**p**rep-**c**oord).

Turkish and Croatian: a closer look

Turkish

- Increase for core arguments: csubj, ccomp
- Decrease for non-core arguments: coordination
- Low size of Turkish treebank, more quality control needed? Results might not be representative...

Croatian

- Chain-like encoding of coordinations improves results for all GFs

$\rightarrow$ keep UD annotations, only change encoding of coordinations: 77.7% $\rightarrow$ 78.6% CNC

Inconsistencies in Turkish annotation of coordinations

1	Ekonomik	ekonomik	ADJ	...	3	conj
2	ve	ve	CONJ	...	3	conj
3	Sosyal	sosyal	ADJ	...	4	amod
4	Konsey'e	Konsey	PROPN	...	6	nmod
5	işlerlik	işlerlik	NOUN	...	6	dobj
6	kazandırılsın	kazan	VERB	...	0	root
7	. .	PUNCT	Punc	...	6	punct

DLM ratio for individual conversions

	Lang	orig DL	cop	prep	coord	c-p-c	Δ (IMS)
ger	de	3.4	0.98	1.01	1.03	1.03	1.3
	en	2.9	1.00	1.04	1.03	1.07	0.8
ira	fa	3.5	0.97	0.99	1.02	0.97	3.7
rom	ca	3.1	1.00	1.06	1.03	1.09	1.6
	es	2.8	0.99	1.07	1.04	1.11	1.9
	fr	2.8	0.99	1.07	1.03	1.09	2.5
	it	2.7	1.00	1.05	1.02	1.08	1.8
	ro	2.7	1.00	1.04	1.04	1.07	0.3
sla	bg	2.5	1.01	1.05	1.02	1.08	1.7
	cs	2.8	1.00	1.58	1.03	1.06	1.2
	hr	2.8	1.00	1.03	1.04	1.08	0.2
	ru	2.7	1.00	1.02	1.03	1.05	1.2
sin	zh	3.6	1.00	0.98	1.01	1.00	1.3
fin	et	2.6	1.00	1.00	1.03	1.02	1.4
tur	tr	2.6	1.00	1.01	1.01	1.02	-0.3

Avg. DL in the original treebank and DLM ratio for each modification

ADE: Results for different conversions

	lang	Δ cop	Δ prep	Δ coord	Δ c-p-c
ger	de	-0.26	-0.03	0.03	-0.23
	en	-0.56	-0.19	-0.01	-0.72
ira	fa	-0.73	0.07	0.02	-0.60
rom	ca	0.09	0.07	-0.01	0.16
	es	-0.19	-0.19	0.02	-0.36
	fr	-0.16	-0.15	0.04	-0.27
	it	-0.22	-0.11	0.02	-0.29
	ro	-0.13	0.17	0.04	0.09
sla	bg	-0.31	-0.10	0.05	-0.34
	cs	-0.30	0.20	0.07	0.03
	hr	0.16	0.21	0.03	0.41
	ru	0.17	0.19	0.05	0.41
sin	zh	-0.25	-0.00	0.03	-0.19
fin	et	-0.37	0.16	0.04	-0.16
tur	tr	0.19	0.28	0.03	0.50

Table: Difference (Δ) between avg. unlexicalised arc direction entropy (ADE) in the original treebank and in the modified treebanks