Tutorial on Universal Dependencies

Making use of UD in NLP and linguistics

Joakim Nivre[1] Daniel Zeman[2] Filip Ginter[3] Francis M. Tyers[45]

[1]Department of Linguistics and Philology, Uppsala University, Sweden

[2]Institute of Formal and Applied Linguistics, Charles University in Prague, Czech Republic

[3]Department of Information Technology, University of Turku, Finland

[4]Giela ja kultuvrra instituhtta, UiT Norgga árktalaš universitehta, Tromsø, Norway

[5]Arvutiteaduse instituut, Tartu Ülikool, Estonia

Parsing

- Monolingual parsing
  - Benchmarking for many different languages
  - Off-the-shelf models:
    - UDPipe
    - SyntaxNet
- Cross-lingual parsing
  - Usually motivated by a low-resource scenario
  - Three main approaches
    - Annotation projection (Hwa et al., 2002)
    - Model transfer (Zeman and Resnik, 2008)
    - Treebank translation (Tiedemann et al., 2014)
- Universal parsing
  - A single model for all languages (Ammar et al., 2016)

Cross-Lingual Parsing

McDonald et al. (2011)

Target Test Language	Source Training Language: da	de	el	en	es	it	nl	pt	sv
da	**79.2**	45.2	44.0	45.9	45.0	48.6	46.1	48.1	47.8
de	34.3	**83.9**	53.2	47.2	45.8	53.4	55.8	55.5	46.2
el	33.3	52.5	**77.5**	63.9	41.6	59.3	57.3	58.6	47.5
en	34.4	37.9	45.7	**82.5**	28.5	38.6	43.7	42.3	43.7
es	38.1	49.4	57.3	53.3	**79.7**	68.4	51.2	66.7	41.4
it	44.8	56.7	66.8	57.7	64.7	**79.3**	57.6	69.1	50.9
nl	38.7	43.7	62.1	60.8	40.9	50.4	**73.6**	58.5	44.2
pt	42.5	52.0	66.6	69.2	68.5	74.7	67.1	**84.6**	52.1
sv	44.5	57.0	57.8	58.3	46.3	53.4	54.5	66.8	**84.8**

Cross-Lingual Parsing

McDonald et al. (2011)

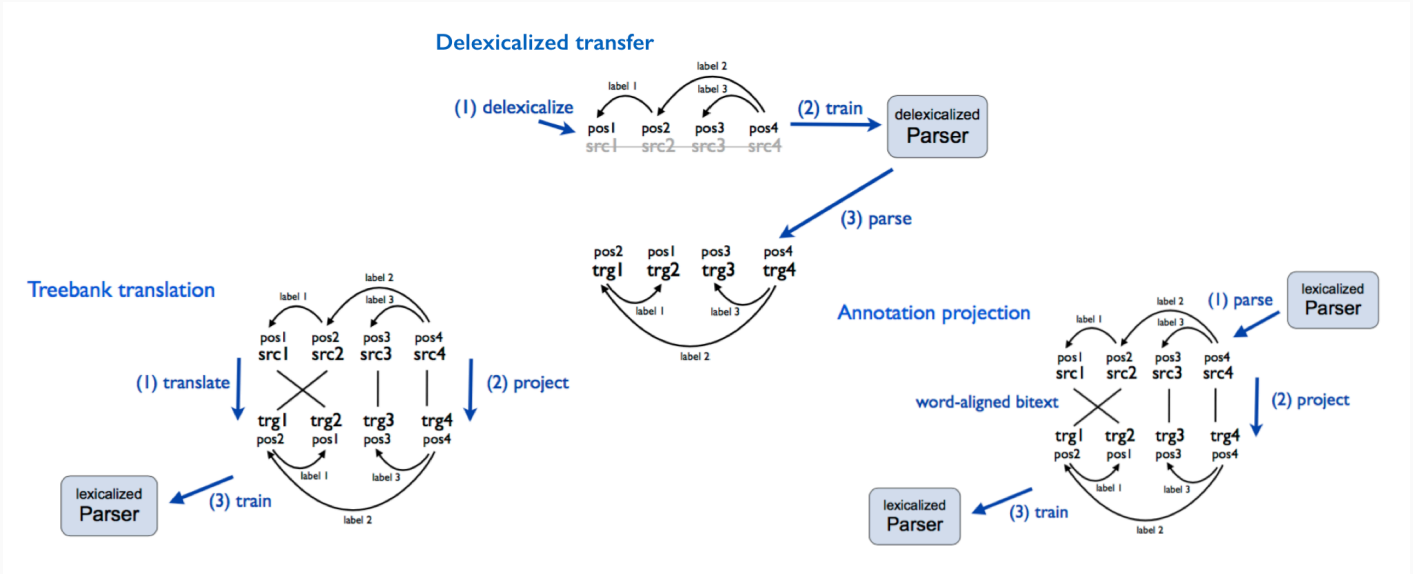
Target Test Language	Source Training Language: da	de	el	en	es	it	nl	pt	sv
da	**79.2**	45.2	44.0	45.9	45.0	48.6	46.1	48.1	47.8
de	34.3	**83.9**	53.2	47.2	45.8	53.4	55.8	55.5	46.2
el	33.3	52.5	**77.5**	63.9	41.6	59.3	57.3	58.6	47.5
en	34.4	37.9	45.7	**82.5**	28.5	38.6	43.7	42.3	43.7
es	38.1	49.4	57.3	53.3	**79.7**	68.4	51.2	66.7	41.4
it	44.8	56.7	66.8	57.7	64.7	**79.3**	57.6	69.1	50.9
nl	38.7	43.7	62.1	60.8	40.9	50.4	**73.6**	58.5	44.2
pt	42.5	52.0	66.6	69.2	68.5	74.7	67.1	**84.6**	52.1
sv	44.5	57.0	57.8	58.3	46.3	53.4	54.5	66.8	**84.8**

McDonald et al. (2013)

Source Training Language	Target Test Language: Unlabeled Attachment Score (UAS)						Labeled Attachment Score (LAS)					
	Germanic			Romance			Germanic			Romance		
	DE	EN	SV	ES	FR	KO	DE	EN	SV	ES	FR	KO
DE	74.86	55.05	65.89	60.65	62.18	40.59	64.84	47.09	53.57	48.14	49.59	**27.73**
EN	58.50	83.33	**70.56**	68.07	70.14	**42.37**	48.11	78.54	**57.04**	56.86	58.20	26.65
SV	**61.25**	**61.20**	80.01	67.50	67.69	36.95	**52.19**	**49.71**	70.90	54.72	54.96	19.64
ES	55.39	58.56	66.84	78.46	**75.12**	30.25	45.52	47.87	53.09	70.29	**63.65**	16.54
FR	55.05	59.02	65.05	**72.30**	81.44	35.79	45.96	47.41	52.25	**62.56**	73.37	20.84
KO	33.04	32.20	27.62	26.91	29.35	71.22	26.36	21.81	18.12	18.63	19.52	55.85

Cross-Lingual Parsing

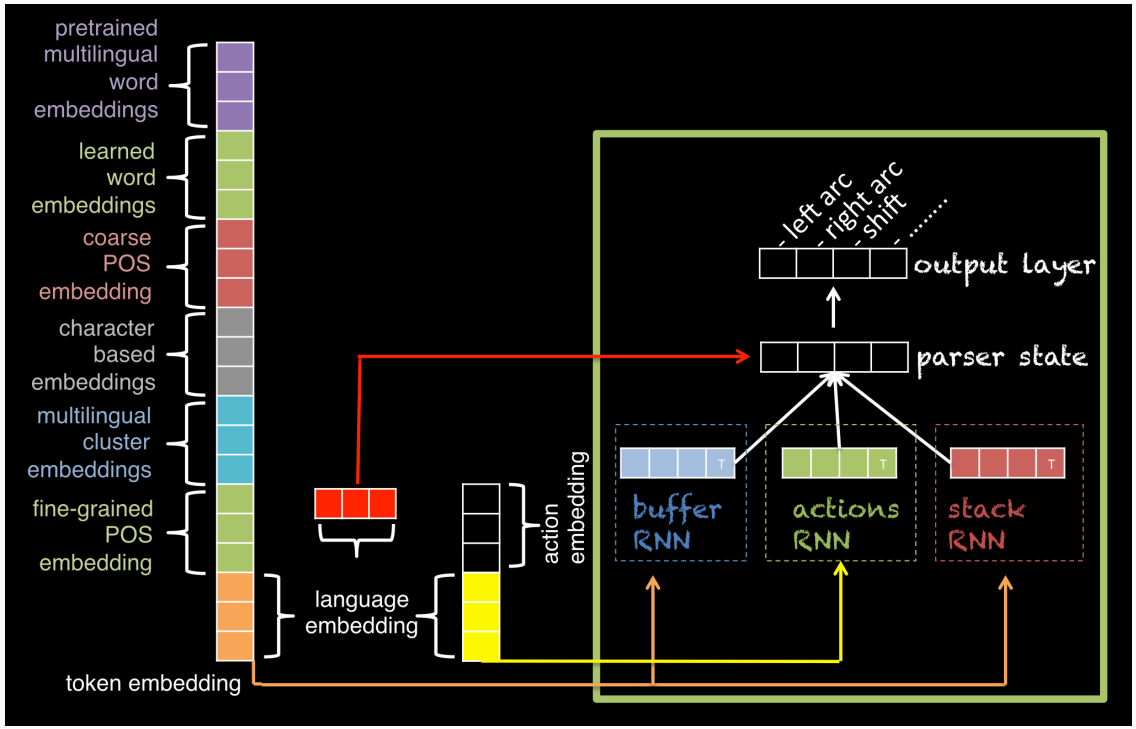
Tiedemann (2015) Cross-Lingual Dependency Parsing
with Universal Dependencies and Predicted PoS Labels



- Three methods for cross-lingual dependency parsing
- The impact of not having gold part-of-speech tags
- Reveals weaknesses of delexicalized model transfer

Universal Parsing

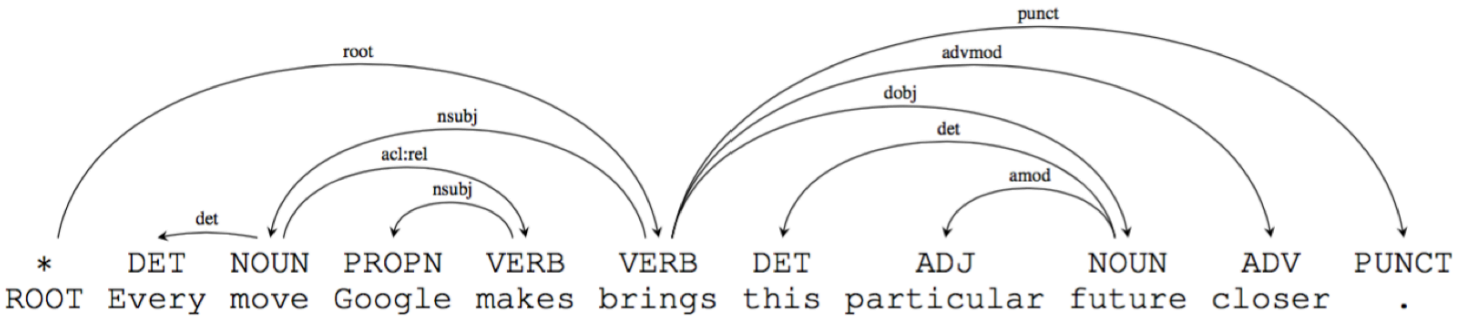
Ammar et al. (2016) Many Languages, One Parser



- Parsing with multiple source and target languages
- Multilingual word embeddings and typological features
- Gain on "small" languages without loss on "big" languages

Need More Data?

Wang and Eisner (2016) The Galactic Dependencies Treebanks: Getting More Data by Synthesizing New Languages



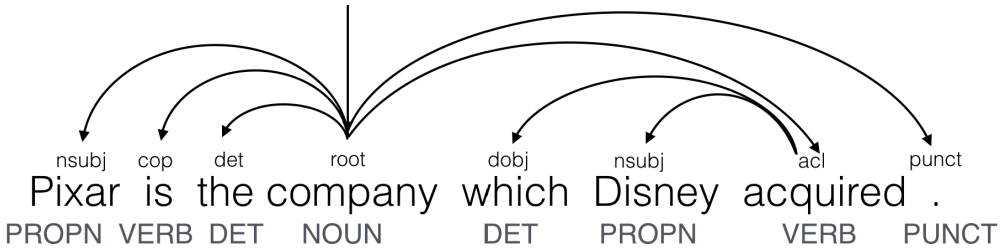
Language	Sentence
English	Every move Google makes brings this particular future closer.
English[French/N]	Every move Google makes brings this future particular closer.
English[Hindi/V]	Every move Google makes this particular future closer brings.
English[French/N, Hindi/V]	Every move Google makes this future particular closer brings.

- Synthesizing treebanks for new (potential) natural languages
- Reorder N- and/or V-dependents in L_1 with model trained on L_2

Semantic Parsing

Reddy et al. (2016) Transforming Dependency Structures to Logical Forms for Semantic Parsing

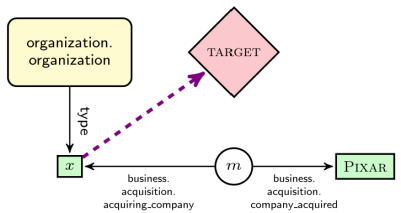
Take syntactic dependencies



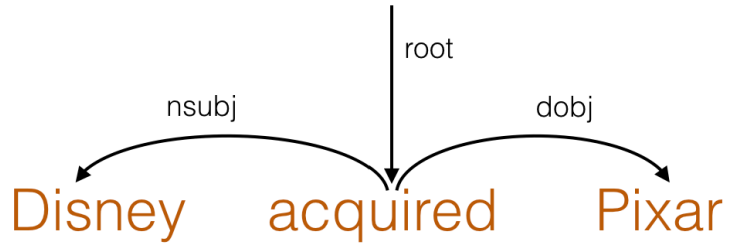
Deterministically infer logical form(s)

$$\exists z.\text{company}(\text{Pixar}) \wedge \text{acquired}(z_e) \wedge \text{arg}_1(z_e, \text{Disney}) \wedge \text{arg}_2(z_e, \text{Pixar})$$

Learn model to map logical form to KG to answer question



Semantic Parsing



binarization

… > dobj > … > nsubj > ….

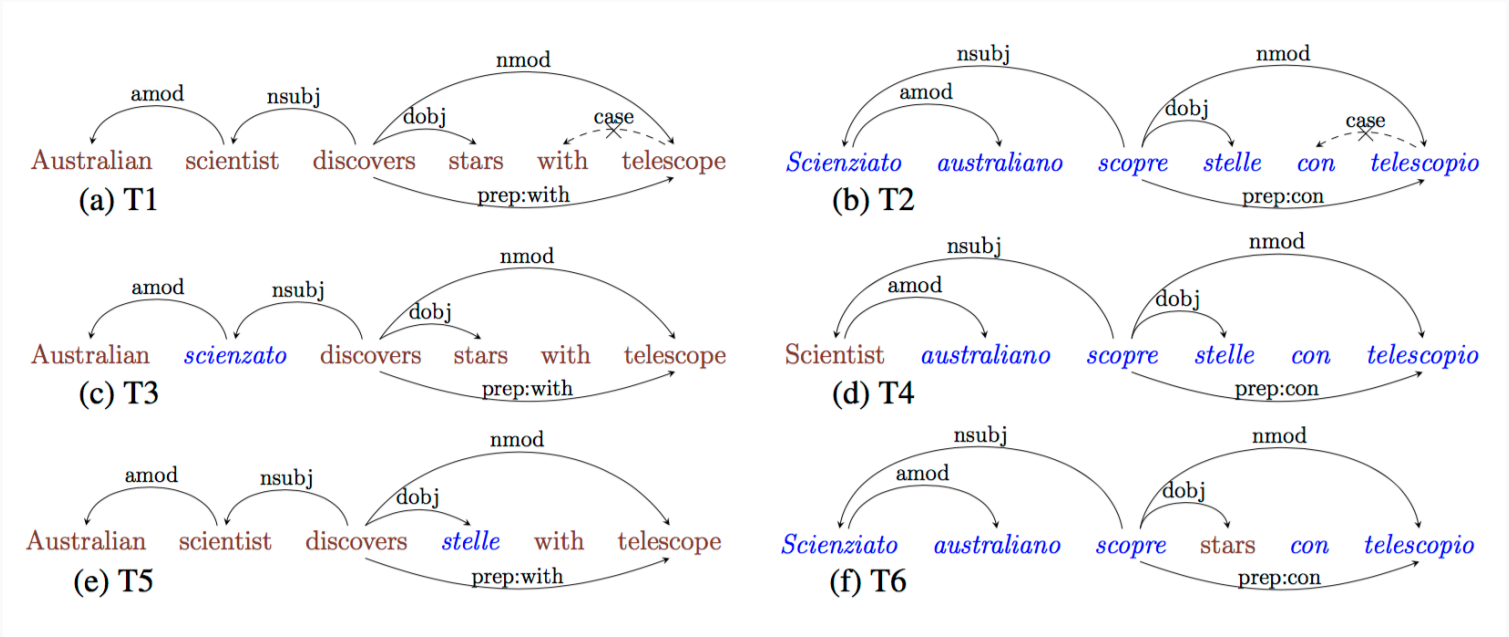
(nsubj (dobj acquired Pixar) Disney)

substitution + composition

$$\lambda z.\exists xy.\text{acquired}(z_e) \wedge \text{Pixar}(y_a) \wedge \text{Disney}(x_a) \wedge \text{arg}_1(z_e, x_a) \wedge \text{arg}_2(z_e, y_a)$$

Cross-Lingual Word Embeddings

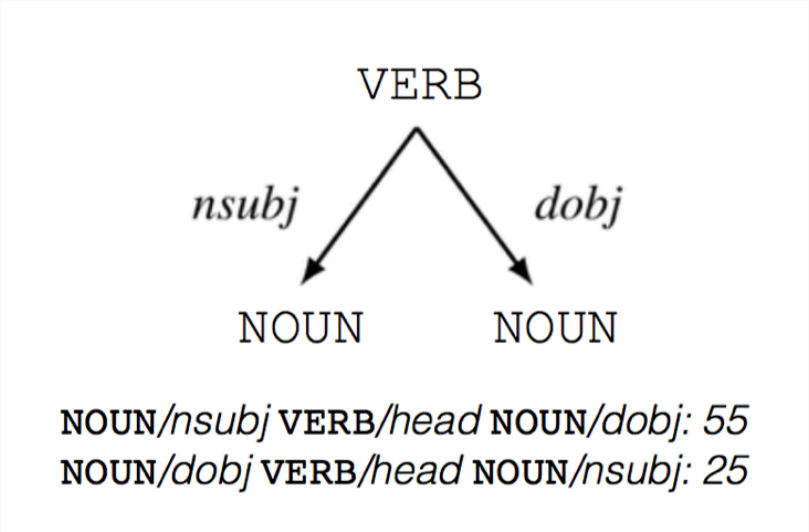
Vulić (2017) Cross-Lingual Syntactically Informed Distributed Word Representations



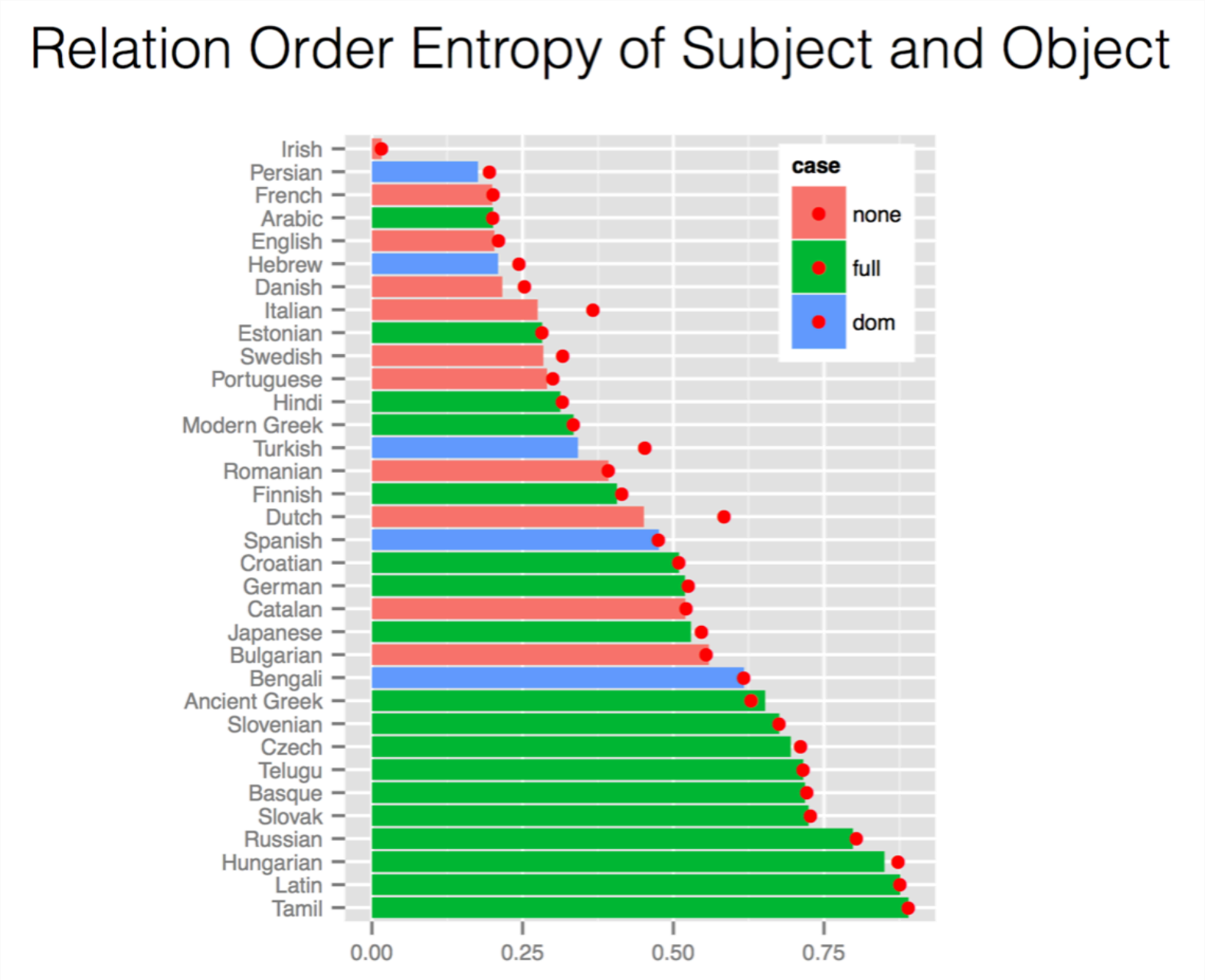
- Cross-lingual UD-parsed corpora (using bilingual lexicon)
- Evaluated on word similarity and bilingual lexicon induction
- Outperforms cross-lingual/no syntax and monolingual/syntax

Linguistic Typology

Futrell et al. (2015) Quantifying Word Order Freedom in Dependency Corpora



- Word order freedom = conditional entropy of order given tree
- Test hypotheses about case marking and word order freedom


Relation Order Entropy of Subject and Object
Irish
Persian
French
Arabic
English
Hebrew
Danish
Italian
Estonian
Swedish
Portuguese
Hindi
Modern Greek
Turkish
Romanian
Finnish
Dutch
Spanish
Croatian
German
Catalan
Japanese
Bulgarian
Bengali
Ancient Greek
Slovenian
Czech
Telugu
Basque
Slovak
Russian
Hungarian
Latin
Tamil
0.00
0.25
0.50
0.75
case
none
full
dom

Linguistic Typology

Östling (2015) Word Order Typology through Multilingual Word Alignment

Feature	Languages	Types	Tokens	Most common
81A: Subject, Object, Verb (Dryer, 2013e)	342	85.4%	85.7%	SOV: 43.3%
82A: Subject, Verb (Dryer, 2013d)	376	89.4%	90.4%	SV: 79.8%
83A: Object, Verb (Dryer, 2013c)	387	96.4%	96.4%	VO: 54.8%
85A: Adposition, Noun Phrase (Dryer, 2013b)	329	94.8%	95.1%	Prep: 50.4%
87A: Adjective, Noun (Dryer, 2013a)	334	85.9%	88.0%	AdjN: 68.9%

- Word order study based on Bible translations (986 languages)
- Massively parallel alignment and UD annotation projection
- Evaluation against WALS for a subset of languages

Is UD really suitable for all it is used for?

Manning's Law

The secret to understanding the design of UD is to realize that it is a very subtle compromise between approximately 6 things:

1. UD needs to be satisfactory for analysis of individual languages.
2. UD needs to be good for linguistic typology.
3. UD must be suitable for rapid, consistent annotation.
4. UD must be suitable for computer parsing with high accuracy.
5. UD must be easily comprehended and used by a non-linguist.
6. UD must provide good support for downstream NLP tasks.

It's easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.

Questions?