

When Annotation Schemes Change Rules Help: A Configurable Approach to Coreference Resolution beyond OntoNotes

AMIR ZELDES & SHUO ZHANG

{amir.zeldes,ssz6}@georgetown.edu



GEORGETOWN UNIVERSITY

CORBON Workshop@NAACL 2016, Dan Diego

Why configurable rules?

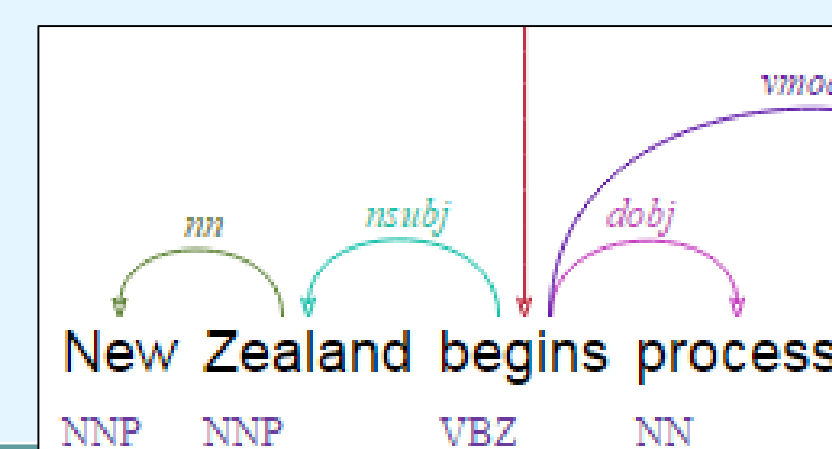
- Annotation schemes differ substantially in included coreference phenomena
- Some examples not annotated in OntoNotes (Hovy et al. 2006):
 - Cataphora/empty pronouns:** *[it]'s certainly true [the rout began immediately after the UAL trading halt]*
 - Predicatives:** *[He] is [an avid fan of a proposition on next week's ballot]*
 - Indefinite/generic:** *[Program trading] is "a racket," ... [program trading] creates deviant swings*
 - Compound modifiers:** *small investors seem to be adapting to greater [stock market] volatility ... Glenn Britta ... says he is "factoring" [the market's] volatility "into in-vestment decisions."*
 - Metonymy:** *a strict interpretation of a policy requires [The U.S.] to notify foreign dictators of certain coup plots ... [Washington] rejected the bid*
 - Nesting:** *He has in tow [his prescient girlfriend, whose sassy retorts mark [her].]*
- Can we make a configurable system that switches these on and off without different training data?

xrenner

- eXternally configurable REference and Non Named Entity Recognizer

- Purely **dependency based** (easier to get data with Universal Dependencies, de Marneffe et al. 2014)

- DepEdit** module manipulates input:
(<http://corpling.uis.georgetown.edu/depedit>)



	JJ-that-cataphora	age-appos
toks	text=/^[li]t\$/&func=/nsubj/; pos=/JJ/; func=/ccomp/	pos=/^NNP\$/; text=/^,\$/; text=/^[1-9][0-9]*\$/
rels	#2>#1;#2>#3	#1.#2.#3;#1>#3
acts	#3:func=cata	#3:func=appos

- Does its own NER (11 types) and entity subtype recognition (54 types)
- Almost **no language specifics** hard wired – **lots** of configurations: (<100)
 - labels for subject, coordination, possession, modification...
 - modifiers tags which must match in coref, no new modifier setting...
 - labels that signal (pro)nominal heads, functions that break the chain

- Lots of lexical data:

Data	Sources
Is-a/has-a matching	DBpedia (Auer et al. 2007)
Geo-names	DBpedia (Auer et al. 2007)
Is-a list	GUM/OntoNotes/PPDB (Gantikevitch et al. 2013)
Has-a list	PTB (Marcus et al. 1993)
Common nouns	GUM, OntoNotes
Antonyms	OntoNotes, WordNet (Fellbaum '98)
Named entities	GUM+OntoNotes+Freebase (Bollacker et al. 2008)

- Stemming for verbal event coreference

Coreference rules

- Cascade of rules also fully configurable (currently 27 for English)

Examples:	Anaphor (1)	Antecedent (2)	Dir	Dist	Propagate
	form="proper"	form="proper"	←	100	none
	text=\$1				
	lemma="one"	form!="proper"&mod=\$1	←	4	→
	text=/(\his her its)/	form!="pronoun"	→	0	←

Case study: OntoNotes and the GUM corpus

We test on OntoNotes/WSJ and the GUM corpus (Zeldes 2016), which has a very different schema and is too small to train on:

	GUM		OntoNotes/WSJ	
	train	test	Train	test
documents	46	8	540	57
tokens	37,758	6,321	322,335	33,306
nominals	11,677	1,933	104,505	13,162
coreference	7,621	1,294	38,587	3,642
- bridging	488	112	--	--
- predicative	71	14	--	--
- cataphora	52	3	--	--
- compound	506	71	--	--

- Switch between two configurations – note just a few differences:
 - Singletons
 - Coordination wrapping
 - Compound modifiers
 - Clausal modifiers in mention
 - Apposition wrapping
 - Indefinite anaphors ...

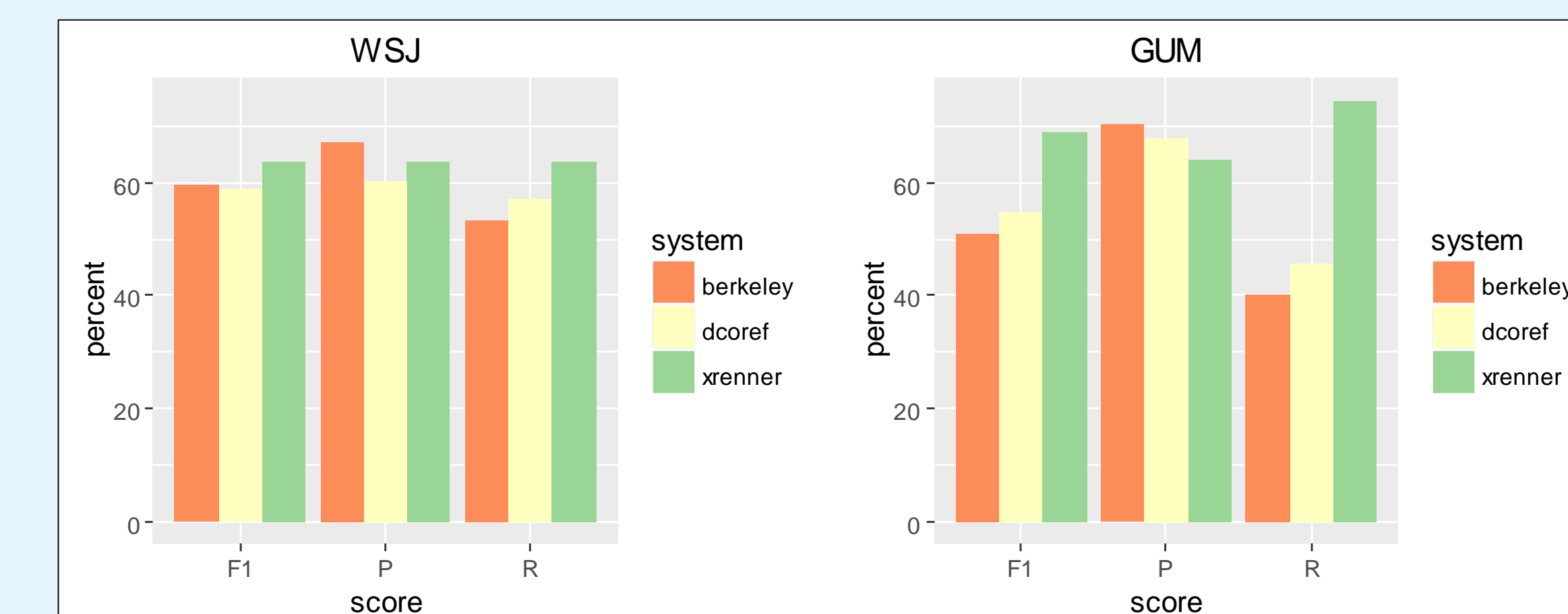
GUM Style

Onto Style

Experiment and results

Experimental setup

- Test standard metrics on both data sets for three systems:
 - CoreNLP dcoref (Lee et al. 2013), rule based system (OntoNotes scheme)
 - Berkeley system (Durrett & Klein 2014), stochastic (trained on OntoNotes)
 - xrenner – two configurations (OntoNotes & GUM style)
- dcoref+Berkeley not trained on GUM scheme – but how big is the difference?
- Auto-parsed** input using Stanford/Berkeley parser (no gold dependencies)
- Mention detection**
 - DepEdit can prevent some errors, raising performance
 - Aggressive coreference matching brings higher recall with small hit to precision
 - GUM scheme reproducible with similar performance using configurations



Coreference resolution

GUM	MUC			B ³			CEAF-e			mean
	R	P	F1	R	P	F1	R	P	F1	F1
xrenner	57.12	54.83	55.95	52.01	46.48	49.09	50.27	39.87	44.47	49.84
dcoref	35.22	57.25	43.61	25.64	50.53	34.02	33.18	39.03	35.87	37.83
berkeley	40.67	71.77	51.92	27.76	60.65	38.09	29.14	52.17	37.40	42.47
WSJ	R	P	F1	R	P	F1	R	P	F1	F1
xrenner	49.47	50.89	50.17	41.13	46.38	43.60	46.17	42.91	44.48	46.08
dcoref	46.77	50.50	48.56	36.41	45.81	40.57	39.93	39.48	39.70	42.94
berkeley	45.07	54.25	49.23	37.30	46.81	41.52	35.21	49.46	41.13	43.96

- Big **is-a** table, **sub-class** based matching (People's Daily > the newspaper)
- Allow **new modifiers** (cf. Lee et al.), use **antonym** check, cardinality
- Problems with overzealous compound matching in GUM (*[[carbon] dioxide]*)

Future work

- Currently working on new language models: German, Hebrew and Coptic
- Dynamic expansion of lexical resources during analysis

References

Auer, S./Bizer, C./Kobilarov, G./Lehmann, J./Ives, Z. 2007. DBpedia: A nucleus for a web of open data. In *6th International Semantic Web Conference*. Busan, South Korea, 11–15.

Bollacker, K./Evans, C./Paritosh, P./Sturge, T./Taylor, J. 2008. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proc ACM SIGMOD 2008*. Vancouver, 1247–1250.

Durrett, G./Klein, D. 2014. A joint model for entity analysis: Coreference, typing, and linking. *TAACL* 2, 477–490.

Fellbaum, C. (ed.) 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Gantikevitch, J./Durme, B. V./Callison-Burch, C. 2013. PPDB: The paraphrase database. In *Proc NAACL-2013*. Atlanta, GA, 758–764.

Hovy, E./Marcus, M./Palmer, M./Ramshaw, L./Weischedel, R. 2006. OntoNotes: The 90% solution. In *Proc NAACL-2006*. New York, 57–60.

Lee, H./Chang, A./Peirsman, Y./Chambers, N./Surdeanu, M./Jurafsky, D. 2013. Deterministic coreference resolution based on entity-centric, precision-ranked rules. *Computational Linguistics* 39(4), 885–916.

Marcus, M. P./Santorini, B./Marcinkiewicz, M. A. 1993. Building a large annotated corpus of English: The Penn treebank. *CL* 19(2), 313–330.

de Marneffe, M.-C./Dozat, T./Silveira, N./Haverinen, K./Ginter, F./Nivre, J./Manning, C. D. 2014. Universal Stanford Dependencies: A cross-linguistic typology. In *Proc. LREC 2014*. Reykjavik, Iceland, 4585–4592.

Zeldes, A. 2016. The GUM corpus: Creating multilayer resources in the classroom. *LRE Journal*.