New supertaggers for the ERG

for the DELPH-IN summit

Olga Zamaraeva & Carlos Gómez-Rodríguez
Department of Informatics/CITIC, Universidade da Coruña

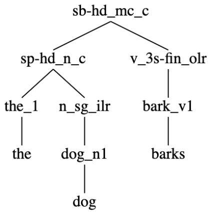
June 28 2023

Introduction
Baseline
Experiments
References

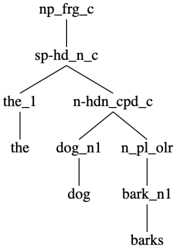
Supertagging helps eliminate unlikely possibilities:

- Without tagging:
  - orthographies are mapped to all possible lexical entries
  - parser has to consider each possibility
  - charts become too big

1



2

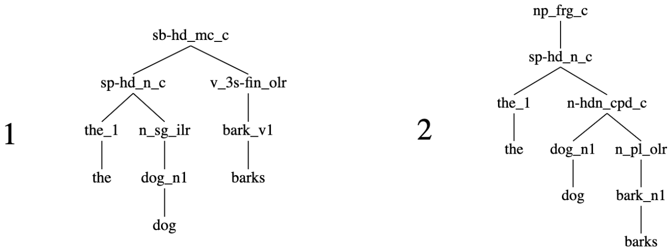


Supertagging is useful for e.g. :

- word sense disambiguation
- parse ranking
- improving parsing speed

Supertagging accuracy is important

- If a wrong lexical type is predicted:
  - the chances of getting the parse right are 0



Prior work on supertagging for HPSG

model	grammar	training tok	tagset size	speed-up factor
N-gram (Prins and van Noord 2004)	Alpino (Dutch)	24 mln	1365	2
HMM (Blunsom 2007)	ERG (English)	113K	615	8.5
MEMM (Dridan 2009)	ERG (English)	158K	676	12

- Dridan (2009):
  - 92% accuracy on in-domain data
  - 74.6% out of domain (up to 80.8% with additional training data)
- Recent work on CCG (Liu et al. 2021):
  - 95.5% accuracy in domain
  - 81% and 92.4% on two out-of-domain datasets

Experiments with ERG

- 2022:
  - Supertagging (no ubertagging)
  - Single tag accuracy (top 1)
- 2023:
  - Started integration into ACE
  - Added fine-tuned BERT

dataset	description	sent	tok	train tok	MaxEnt	SVM	NCRF++	BERT	D2009
cb	technical essay	713	17,244	0	88.96	89.53	91.94	**93.88**	74.61
ecpr	e-commerce	1088	11,550	24,934	91.80	91.99	95.09	**96.09**	
jh*,tg*,ps*, ron*	travel brochures	2116	34,098	147,166	90.45	91.21	95.44	**96.11**	91.47
petet	textual entailment	581	7135	1578	92.88	95.31	96.93	**97.71**	
vm32	phone conv.	1000	8730	86,630	93.57	94.29	95.62	**96.64**	
ws213-214	Wikipedia	1470	29,697	161,623	91.31	92.02	93.66	**95.59**	
wsj23	Wall Street J.	950	22,987	959,709	94.27	94.72	96.05	**97.26**	
all	all test sets as one	7,918	131,441	1,381,645	91.57	92.28	94.46	**96.02**	
all	average	7,918	131,441	1,381,645	91.89	92.72	94.96	**96.18**	
speed (sen/sec)	average	7,918	131,441	1,381,645	1024	**7414**	125	346	

Preliminary Spanish experiments

- Spanish treebanks are much smaller and ‘gold‘ is of lower quality
- Accuracy of the NCRF++ supertagger in the 70%s
  - BERT will be better but probably not much
- Will try to do some multilingual training
  - ...and also improve the Spanish treebanks

Integration into ACE

- For now:
  - Oracle-style experiments for effects on parsing speed
    - ACE gets list of supertags for each sentence
    - eliminates edges with wrong lexical type
- Not sure for now how to integrate the model itself for new input

Example (not a fair comparison!)

No tagging:

```
SENT: Abrams heard barks.

NOTE: 5 readings, added 1194 / 315 edges to chart (124 fully instantiated, 89 actives used, 92 passives used)    RAM: 3975k

NOTE: parsed 1 / 1 sentences, avg 3975k, time 0.02536s
```

Ubertagging:

```
NOTE: 3 readings, added 1133 / 258 edges to chart (101 fully instantiated, 83 actives used, 73 passives used)    RAM: 3661k

NOTE: parsed 1 / 1 sentences, avg 3661k, time 0.02434s
```

BERT oracle supertags:

```
NOTE: 3 readings, added 1052 / 177 edges to chart (73 fully instantiated, 78 actives used, 50 passives used)    RAM: 3413k

NOTE: parsed 1 / 1 sentences, avg 3413k, time 0.02154s
```

Summary

- Some progress on supertagging since last year
  - added BERT
  - started integrating with ACE
  - tried training a supertagger for Spanish

Summary

- Some progress on supertagging since last year
  - added BERT
  - started integrating with ACE
  - tried training a supertagger for Spanish
- BERT is more accurate than other things

Summary

- Some progress on supertagging since last year
  - added BERT
  - started integrating with ACE
  - tried training a supertagger for Spanish
- BERT is more accurate than other things
- scikit-learn SVM (Pedregosa et al. 2011) remains the fastest model

Summary

- Some progress on supertagging since last year
  - added BERT
  - started integrating with ACE
  - tried training a supertagger for Spanish
- BERT is more accurate than other things
- scikit-learn SVM (Pedregosa et al. 2011) remains the fastest model
- ACE integration underway but for now only for treebank experiments

References

Philip Blunsom. 2007. *Structured classification for multilingual natural language processing*. PhD thesis, University of Melbourne.

Rebecca Dridan. 2009. *Using lexical statistics to improve HPSG parsing*. PhD thesis, University of Saarland.

Yufang Liu, Tao Ji, Yuanbin Wu, and Man Lan. 2021. Generating ccg categories. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13443–13451.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

RP Prins and GJM van Noord. 2004. Reinforcing parser preferences through tagging. *Traitement Automatique des Langues*, 3:121–139.