



EVALITA 2011

Evaluation of NLP and Speech Tools for Italian

EVALITA 2011

Domain Adaptation for Dependency Parsing

F. Dell'Orletta, S. Marchi, S. Montemagni, G. Venturi
T. Agnoloni, E. Francesconi



**Istituto di Linguistica
Computazionale**

C. N. R.





Introduction: the motivation

Why

- **Domain Adaptation (DA) at Evalita?**
 - A challenge for all NLP tasks
 - See the “Domain Adaptation Track” at CoNLL 2007 and the workshop “Domain Adaptation for Natural Language Processing 2010” (DANLP 2010)
 - So far, focus on English: first initiative for the Italian language
- **DA for dependency parsing?**
 - A number of different text processing tasks benefit significantly from natural language dependency parsing
- **DA wrt the legal domain?**
 - Linguistically peculiar nature of legal texts wrt open-domain texts
 - Growing demand for real word applications for the legal domain
 - The first NLP shared task devoted to the legal domain



Introduction: the goal

- Investigating techniques for adapting state-of-the-art dependency parsing systems to domains outside of the data from which they were trained or developed
- DA at EVALITA organized into two subtasks:
 - **minimally supervised domain adaptation** with limited annotated resources in the target domain and unlabeled corpora
 - **unsupervised domain adaptation** with no annotated resources in the target domain, i.e. using only unlabeled target data
 - similar to the DA track at CoNLL 2007 with a main difference: here, the target domain of the development and test sets is the same



Dataset: description

- **Source domain**

- Revised version of ISST-TANL corpus (newswire open-domain corpus)

Training set	71,568 tokens (3,275 sentences)	Avg sentence length 21.85 (23.71) tokens
Development set	5,165 tokens (231 sentences)	Avg sentence length 22.36 (24.13) tokens

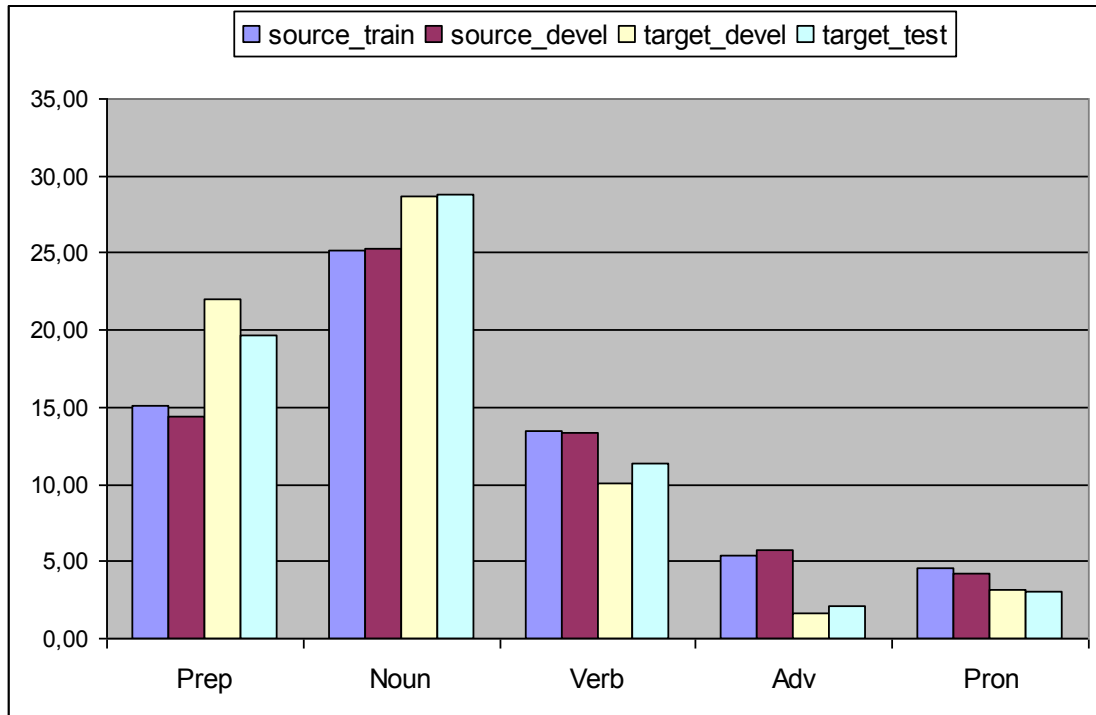
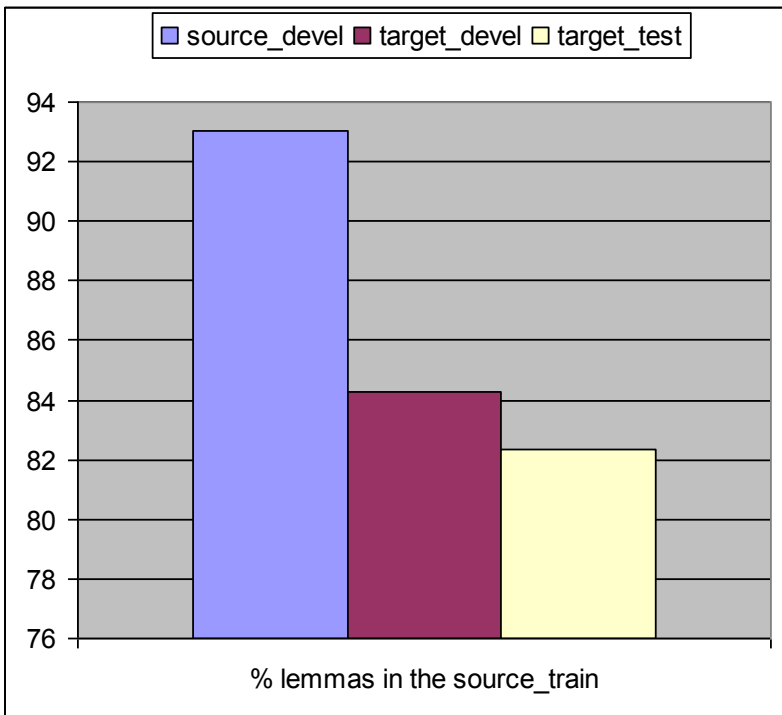
- **Target domain**

- Italian legislative corpus, gathering laws enacted by different releasing agencies (European Commission, Italian State and Regions)

Manually annotated development set	5,664 tokens (147 sentences)	Avg sentence length 38.53 (38.77) tokens
Automatically PoS tagged corpus	13,010,610 tokens (620,064 sentences)	Avg sentence length 20.98 (31.46) tokens
Manually annotated test set used for the evaluation	5,358 tokens (170 sentences)	Avg sentence length 31.52 (37.84) tokens



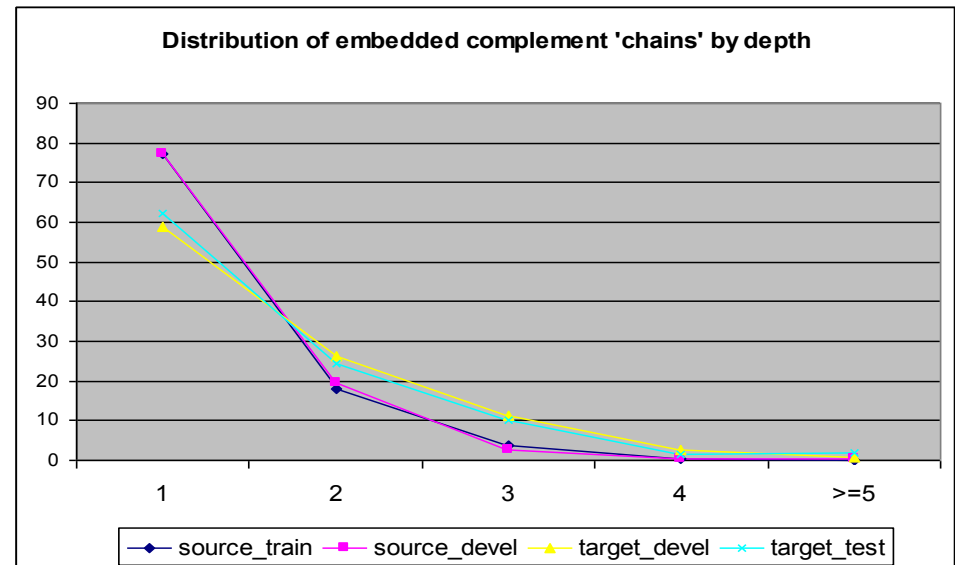
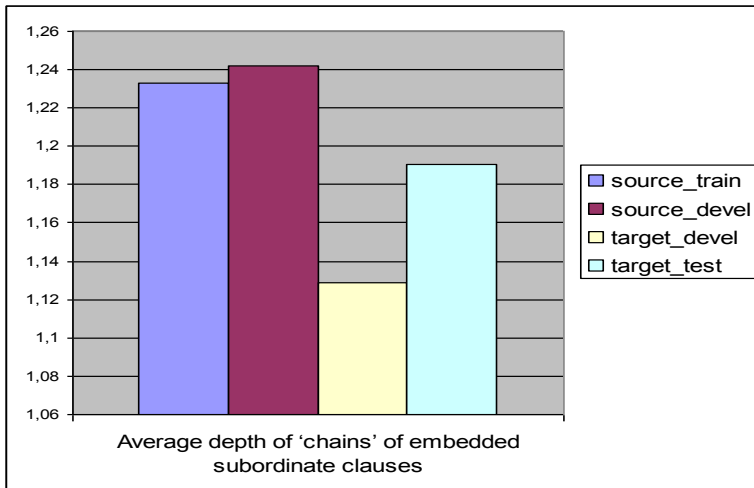
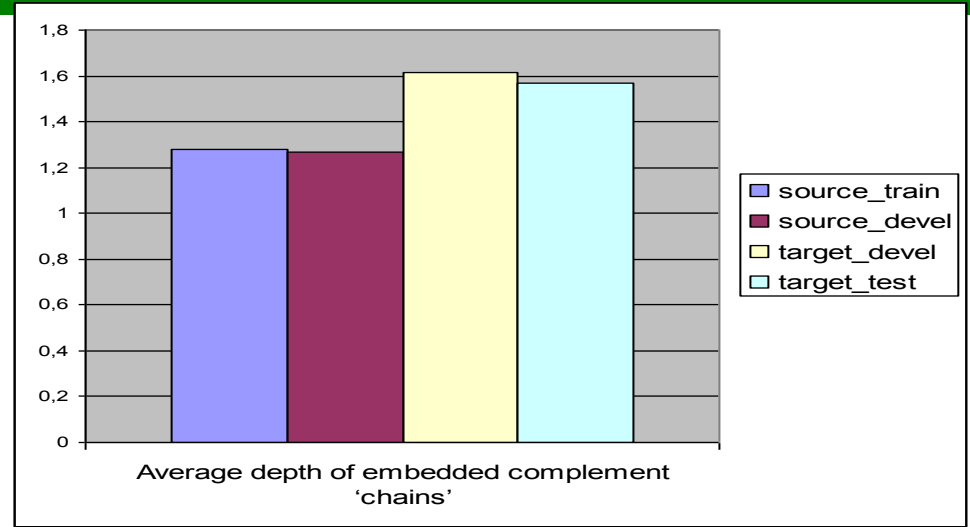
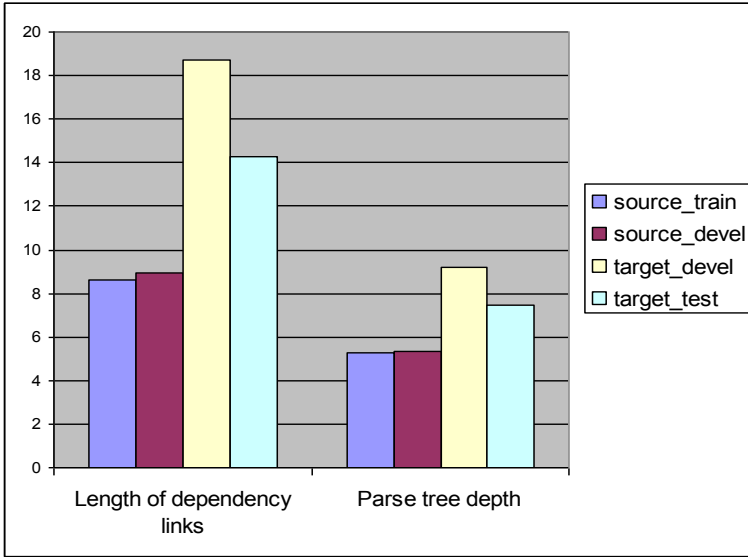
Dataset: lexical and morpho-syntactic features



Lexical level: some differences exist but not so deep if compared wrt other target domains
Morpho-syntactic level: significant differences for many POS (e.g. +5% prep)



Dataset: syntactic features





Systems' results: participants

- Group 1:
 - Giuseppe Attardi, Maria Simi, Andrea Zanelli (University of Pisa)
 - Task: minimally supervised domain adaptation and unsupervised domain adaptation
 - Base parsing model: DeSR parser, a Shift/Reduce deterministic transition-based parser
 - DA method: Active learning
- Group 2:
 - Barbara Plank (University of Trento) and Anders Søgaard (University of Copenhagen)
 - Task: unsupervised domain adaptation
 - Base parsing model: the second-order projective model of MST parser, a graph-based parser
 - DA method: two self-training methods



Systems' results: LAS

- Minimally supervised domain adaptation subtask

System	Source Devel	Test
Attardi et al. (Base[a])	82.09	80.29
Attardi et al. (DA)	82.34	81.39

+1%

- Unsupervised domain adaptation subtask

System	Source Devel	Test
Attardi et al. (Base[b])	82.09	75.85
Attardi et al. (DA)	81.09	80.83
Plank&Søgaard (Base)	80.19	74.62
Plank&Søgaard (DA1)	80.87	74.02
Plank&Søgaard (DA2)	80.31	74.30

+5%

No improv



Discussion

- Two dependency parsers and two different approaches to DA
 - Not comparable results
- Summing up:
 - good performance (+5%) of the active learning strategy
 - no improvement achieved with self-training approaches
- Possibly due to the syntactic peculiarities characterizing legal texts
 - The formulaic nature of legal language can help explaining
 - why a small amount of new target data are enough to enable the parser to reliably handle new syntactic structures specific to the target domain
 - why self-training approaches could not be reliable (the “strange” case of the target development set)



Conclusion

- EVALITA DA results are in contrast with previous self-training experiments reported in the literature where good results are achieved wrt different target domains (e.g. biomedical, chemical)
 - Biomedical/chemical vs newswire texts: the main differences hold at the level of lexical features
 - Legal vs newswire texts: the main differences are concerned with the syntactic structure of the text
- This opens new avenues for DA:
 - based on different DA approaches/strategies
 - dealing with different languages and target domains
- Post-EVALITA event: First Shared Task on Dependency Parsing of Legal Texts will be organized within the LREC 2012 Workshop “Semantic Processing of Legal Texts” (http://poesix1.ilc.cnr.it/splet_shared_task/)
 - The languages dealt with will be English and Italian