

Hybrid Multilingual Parsing with HPSG for Semantic Role Labeling

Rui Wang
rwang@coli.uni-sb.de



Saarland University
Germany

Yi Zhang
yzhang@coli.uni-sb.de



DFKI GmbH, Germany

Stephan Oepen
oe@ifi.uio.no



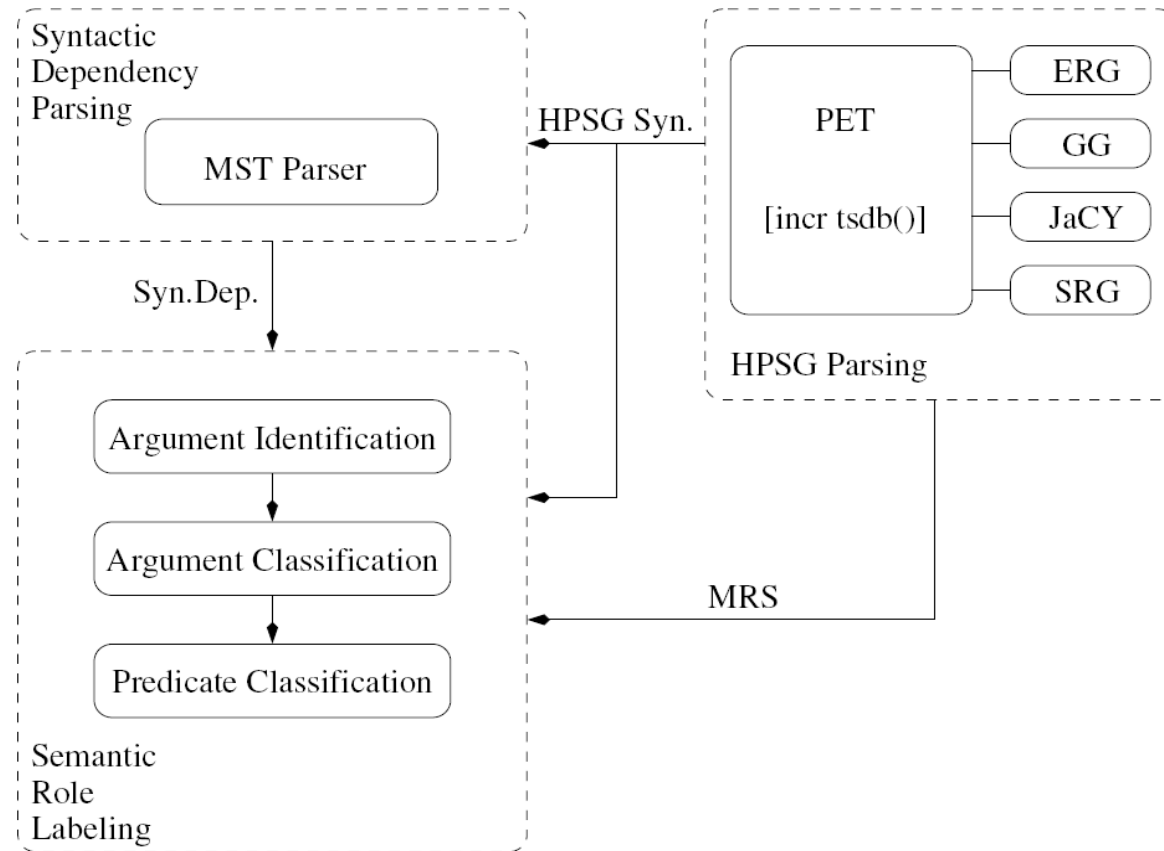
University of Oslo
Norway



Motivations

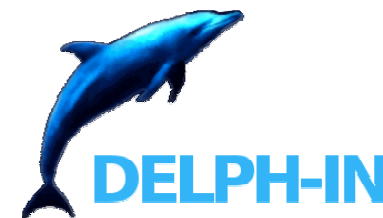
- The extension of Zhang et al. (2008)
 - From English to multiple languages
 - Both syntactic and semantic features
- Road-test the hand-written DELPH-IN HPSG grammars being developed in the past two decades

Architecture



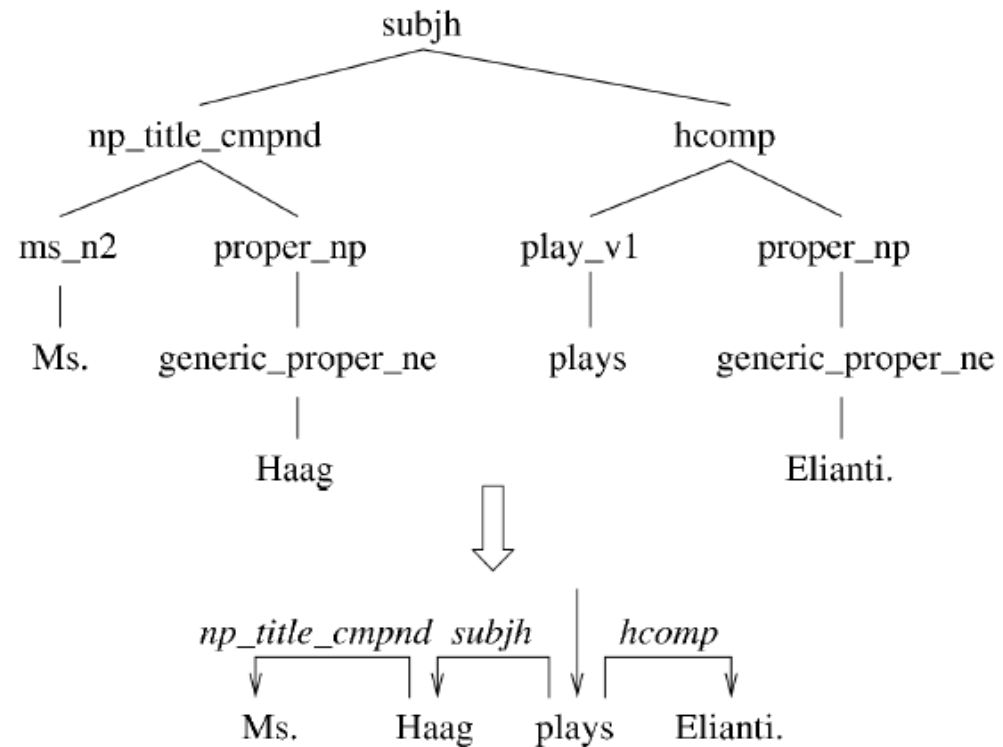
Deep Linguistic Grammars

- ERG (en)
 - Dan Flickinger
 - Coverage: 80.4%
- GG (de)
 - Berthold Crysmann, Peter Adolphs
 - Coverage: 28.6%
- JaCY (ja)
 - Francis Bond
 - Coverage: 42.7%
- SRG (es)
 - Montserrat Marimon
 - Coverage: 7.5%



<http://www.delph-in.net/>

Dependency Backbone Extraction



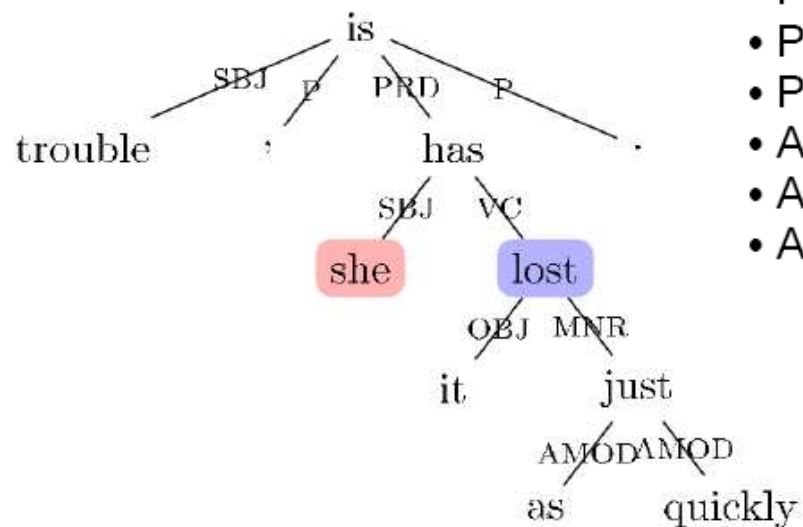


HPSG Parsing

- PET: Efficient HPSG Parser
 - Chart-mapping-based re-tokenization
 - Unknown word handling with POS mapping rules
 - Efficient best-first parsing with ambiguity packing
- Retraining Parse Disambiguation Models
 - Original models trained with manually disambiguated HPSG treebanks
 - Retrained to maximize the agreement between HPSG dependency backbone and CoNLL unlabeled syntactic dependencies

MRS Features

[pron_rel			[pronoun_q_rel			[_lose_v_1_rel		
LBL	h3	h	LBL	h5	h	LBL	h8	h
ARG0	x4	x	ARG0	x4		ARG0	e2	
			RSTR	h6	h	ARG1	x4	
			BODY	h7	h	ARG2	x9	x



- P MRS ep-name: lose_v_1_rel
- P MRS-args labels: ARG1 ARG2
- P MRS-args POSEs: PRP PRP
- A MRS ep-name: pron_rel
- A MRS-preds labels: ARG1
- A MRS-preds POSEs: VBZ



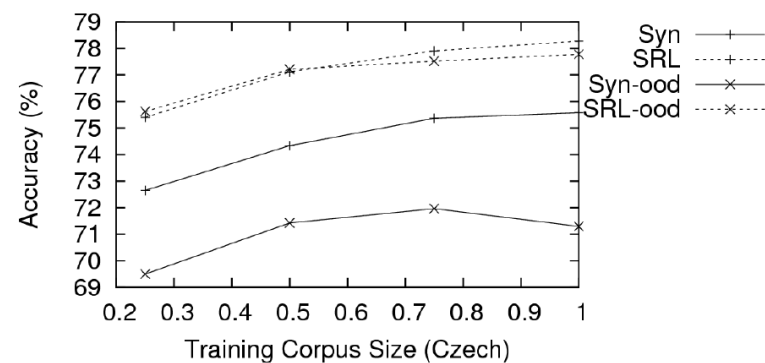
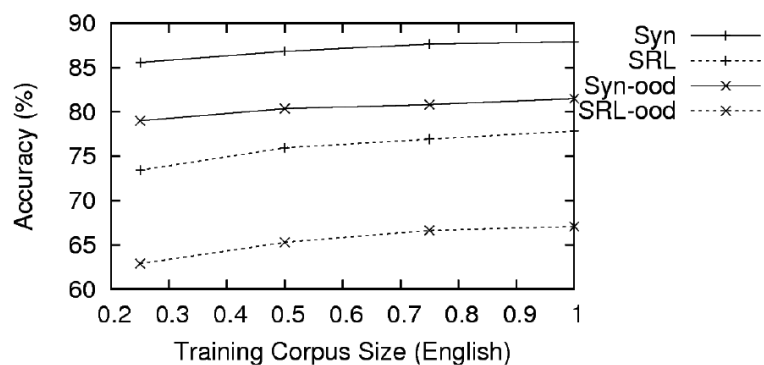
Experiment Settings

- Syntactic Parsing
 - MST Parser
- Semantic Role Labeling
 - MaxEnt-based pipeline of classifiers, with parameters estimated with the open source TADM system; open systems include extra MRS and HPSG DB features
- Parameter tuning

		ca	zh	cs	en	de	ja	es
Syn	proj	+	+	-	+	+	-	+
	morph	-	N/A	+	N/A	-	-	-
SRL	closed	+	+	+	+	+	+	+
	open	-	-	-	+	+	+	+

Results

		ca	zh	cs	en	de	ja	es
SYN	Closed	82.67	73.63	75.58	87.90	84.57	91.47	82.69
	ood	-	-	71.29	81.50	75.06	-	-
SRL	Closed	67.34	73.20	78.28	77.85	62.95	64.71	67.81
	ood	-	-	77.78	67.07	54.87	-	-
	Open	-	-	-	78.13 (↑0.28)	64.31 (↑1.36)	65.95 (↑1.24)	68.24 (↑0.43)
	ood	-	-	-	68.11 (↑1.04)	58.42 (↑3.55)	-	-





Conclusion and Future Work

- The results clearly show that the integration of HPSG parsing results in the semantic role labeling task brings substantial performance improvement, even where grammar coverage is low
- The gain is more significant on out-of-domain tests, indicating that the hybrid system is more robust to cross-domain variation
- The closed SRL system needs to be improved in the future

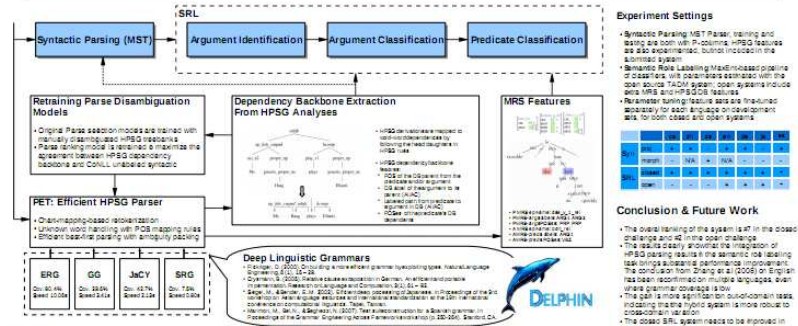
Thank You!

Welcome to our poster!

Hybrid Multilingual Parsing with HPSG for Semantic Role Labeling

Yi Zhang  yzhang@coli.uni-sb.de DFKI GmbH, Germany
 Rui Wang  wang@coli.uni-sb.de Saarland University, Germany
 Stephan Oepen  oe@ifi.uio.no University of Oslo, Norway

Abstract
 In this paper we present our syntactic and semantic dependency parsing system, which extends the system of Zhang, Wang, & Ljckfors (2008) in the multilingual direction, and achieves 76.49 average macro F1 Score on the closed syntactic task. Substantial improvements to the open SRL task have been observed that are attributed to the HPSG parses with handcrafted grammars.



Experiment Settings

- Syntactic Parsing (MST) tasks: training and testing are both with POS-tokens. HPSG features are also experimented, but not included in the automatic system.
- Semantic Role Labeling (MRS): closed parsing of class files with open classes extracted with the open source TACU system; open systems include some MRS and HPSG features.
- Parameter tuning features are the manual selection for each approach on development sets for both closed and open systems.

System	ERG	GG	JCY	SRG
Baseline	75.50	77.85	62.95	64.31
Proposed	76.49	77.85	62.95	64.31

Conclusion & Future Work

- The overall ranking of the system is 47 in the closed challenge and 49 in the open challenge.
- The results clearly showed that the integration of HPSG parsing helps in the semantic role labeling task on long dependency performance improvement. The conclusion from Zhang et al. (2008) on English has been confirmed on multiple languages with more significant coverage in text.
- The gain in the application for multilingual tasks indicates that the hybrid system is more robust to cross-domain adaptation.
- The closed SRL system needs to be improved in the future.

Results

	ca	ab	es	de	ja	cs
Closed	82.67	79.63	75.50	72.90	64.22	62.61
Open	-	77.28	64.30	73.86	-	-
Closed	87.04	78.89	76.28	77.85	62.95	64.31
Open	-	77.78	67.67	74.67	-	-
Open	-	-	76.13 (10.98)	64.81 (11.36)	65.93 (11.24)	64.24 (10.43)
Open	-	-	68.11 (11.01)	56.42 (13.55)	-	-



Acknowledgments

The first author thanks the German Excellence Cluster of Multimodal Computing and Interaction for its support of the work. The second author is funded by the PIRE PhD scholarship program. The third author is supported by the University of Oslo as part of its research partnership with the Centre for the Study of Language and Information at Stanford University. Our deep parsing experimentation was evaluated on the TITAN HPC facilities at the University of Oslo.

Reference

• Oepen, S., Oepen, S., Cramer, U., Czymara, B., Flöbge, D., & Keller, S. (2008). Some tripoints of hybrid natural language parsing. In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC08), Marrakech, Morocco.
 • Hanzl, R. (2002). A comparison of algorithms for maximum entropy semantic extraction. In Proceedings of the 8th conference on natural language learning (CONLL 2002), Taipei, Taiwan.
 • McDonald, R., Forcada, P., de Mar, F., and Ji, H. (2008). Non-invasive pipeline for parsing with Semantic Role Extraction. In Proceedings of NLP 2008, Vancouver, Canada.
 • Zhang, Y., Wang, R., & Ljckfors, N. (2008). Hybrid Learning of Dependency Structures from Multilingual Linguistic Resources. In Proceedings of the 7th International Conference on Computational Natural Language Learning (CONLL 2008), Manchester, UK.