Universal Dependency Parsing with a General Transition-Based DAG Parser

Daniel Hershcovich^{1,2} & **Omri Abend**² & **Ari Rappoport**²

¹The Edmond and Lily Safra Center for Brain Sciences
²School of Computer Science and Engineering
The Hebrew University of Jerusalem {danie

{danielh,oabend,arir}@cs.huji.ac.il



האוניברסיטה העברית בירושלים THE HEBREW UNIVERSITY OF JERUSALEM

ביה״ם להנדסה ולמדעי המחשב ע״ש רחל וסלים בנין The Rachel and Selim Benin School of Computer Science and Engineering

Learning to parse enhanced dependencies jointly with basic Universal Dependency Parsing.

github.com/CoNLL-UD-2018/HUJI

We extend TUPA [2, 3], a general DAG parser originally designed for UCCA: transition-based
parser supporting reentrancy (DAG), discontinuity (non-projectivity) and non-terminal nodes.
Transitions:

Shift, Reduce, Node_X, Right-Edge_X, Left-Enhanced_X, Right-Enhanced_X, Swap, Finish

Parse	Parser state:						
ck			mada	fer	to feel very w		

Transition classifier: transition

Enhanced Dependencies

Some UD treebanks contain enhanced graphs with additional or augmented edges [5, 4].

Conjoined predicates and arguments:





Unified DAG Format

We convert UD into a UCCA-like format supported by TUPA, by inserting non-terminal nodes.



UCCA (Universal Conceptual Cognitive Annotation): cross-lingual semantic representation [1]. Nodes are scenes/concepts. *Primary edges* form a tree. *Remote edges* (dashed) allow reentrancy.



Results

nsubj:xsubj

nsubj:xsubj

	TUPA	TUPA	UDPipe
	(official)	(unofficial)	(baseline)
All treebanks	53.69	58.48	65.80
Big treebanks	62.07	67.36	74.14
PUD treebanks	56.35	56.82	66.63
Small treebanks	36.74	41.19	55.01
Low-resource	8.53	12.68	17.17

Macro-averaged LAS-F1 on test treebanks. (*Unofficial*: after some bug fixes.)

TUPA: first general parser for enhanced UD.

References

[1] Omri Abend and Ari Rappoport. Universal Conceptual Cognitive Annotation (UCCA). In *Proc. of ACL*, pages 228–238, August 2013.

- [2] Daniel Hershcovich, Omri Abend, and Ari Rappoport. A transition-based directed acyclic graph parser for UCCA. In *Proc. of ACL*, pages 1127–1138, 2017.
- [3] Daniel Hershcovich, Omri Abend, and Ari Rappoport. Multitask parsing across semantic representations. In *Proc. of ACL*, pages 373–385, 2018.
- [4] Siva Reddy, Oscar Täckström, Slav Petrov, Mark Steedman, and Mirella Lapata. Universal semantic parsing. In *Proc. of EMNLP*, pages 89–101, 2017.
- [5] Sebastian Schuster and Christopher D. Manning. Enhanced English Universal Dependencies: An improved representation for natural language understanding tasks. In *Proc. of LREC*. ELRA, May 2016.

	LAS-F1	Enhanced LAS-F1
TUPA (unofficial)	72.10	57.13
-NER	71.82	54.65
-POS	69.23	49.12
–Embed.	72.33	54.54
–Remote	72.08	0.00
UDPipe	77.62	0.00
UDPipe + CoreNLP	76.66	21.68

nsubj:pass

obl

Ablation + baselines on **English EWT** dev. (*CoreNLP*: English-specific rule-based postprocessor for enhanced dependencies.)

Please join SemEval 2019 Task 1: Cross-lingual Semantic Parsing with UCCA



tinyurl.com/ semeval-ucca



52.97 52.97 6 20.69 22.78 56.8	8.42 43.27 26.85 34.48 54.48 50.49 50.49 54.13 54.13	35.24 66 48.58 54.04 54.04 50.05 50.75 500	13.01 30.62 52.97 8.99 23.54 31.46 35.13 35.13 35.13 48.92 48.92 48.92 10.23 40.13 40.13 40.13 40.23 43.93 43.93 43.61 36.55 40.13 40.13 43.93 43.93 43.61 40.13 43.61 40.13 43.61 40.13 43.61 40.13 43.61 40.13 43.61 40.13 43.61 40.13 43.61 40.13 43.61 40.13 43.61 43.61 44.20 44.20 56.55 100 100 100 100 100 100 100 100 100 1
af_afribooms grc_perseus grc_proiel ar_padt hy_armtdp eu_bdt br_keb	bxr_bdt bxr_bdt ca_ancora hr_set ca_ancora hr_set cs_fictree cs_pdt cs_pdd da_ddt da_ddt da_ddt nl_alpino nl_alssysmall en_ewt en_ewt en_ewt en_pud en_pud et_edt fo_oft fi ftb	fi_fi_tdt fi_fi_tdt fr_gsd fr_sequoia fr_sequoia fr_spoken gl_treegal de_gsd gl_treegal he_htb he_htb he_htb he_htb he_htb he_btd he_btd he_btd he_btd he_btd he_btd he_btd he_btd he_btd he_btd he_ssd ga_idt it_ja_gsd ja_modern kk_ktb ktb kk_ktb kkb kb kkb kb kb kkb kb kb kb kkb kb kb kb kb kb kkb kb	kmr_mg kmr_mg la_ittb la_ittb la_proiel la_proiel la_proiel lv_lvtb pcm_nsc sme_giella pcm_nsc pr_ssi lfa_seraji pl_lfg pl_sz pl_sz pl_sz sr_set sr_s

LAS-F1 and Enhanced LAS-F1 for TUPA on test treebanks (unofficial).