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	er sta	te:			
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)	
ll treebanks	53.69	58.48	65.80	
Sig treebanks	62.07	67.36	74.14	
UD treebanks	56.35	56.82	66.63	
mall treebanks	36.74	41.19	55.01	
ow-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	20.69	56.8	10.18	8.42			26.85	34.48		53.3	50.49			UT.TU	23.39		35.24				48.58	57.		54.04	50.67	50.75	53.5		Ŭ	10.04	12.12	58	13.01		30.62	23.54	8.99	04.16		35.13			56.5			48.92		27.36	40.13		42.29	46.42 0.23	43.93	15.01		45.67 26 F	C.OS
af_afribooms	grc_perseus	ar_padt	hy_armtdp	br_keb	bg_btb bxr bdt	ca_ancora	hr_set cs_cac	cs_fictree	cs_pdt cs_pud	da_ddt	nl_alpino nl_lassysmall	en_ewt	en_gum	en_lines en pud	et_edt	fo_oft	fi_ftb fi nud	fi_tdt	fr_gsd	fr_sequoia fr_spoken	gl_ctg	gl_treegal	ae_gsa got_proiel	el_gdt	he_htb hi hdtb	hu_szeged	zh_gsd	id_gsd ga_idt	it_isdt	it_postwita	ja_gsd ia modern	kk_ktb	ko_gsd	ko_kalst kmr_mg	la_ittb	la_perseus la proiel	lv_lvtb	pcm_nsc	sme_giella no bokmaal	no_nynorsk	no_nynorsklia fro_srcmf	cu_proiel	fa_seraji n1 Ifo	pl_sz	pt_bosque	ru_syntagrus	ru_taiga	sr_set ob_onk	SK_SIIK Sl_SSj	sl_sst	es_ancora sv lines	_ sv_pud	sv_talbanken th pud	tr_imst	uk_iu hsb_ufal	ur_udtb	ug_udt · · ·	vi_vtb

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