

Universal Dependency Parsing with a General Transition-Based DAG Parser

Daniel Hershcovich^{1,2} & Omri Abend² & Ari Rappoport²

¹The Edmond and Lily Safra Center for Brain Sciences

²School of Computer Science and Engineering

The Hebrew University of Jerusalem

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



האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

ביה"ס 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.

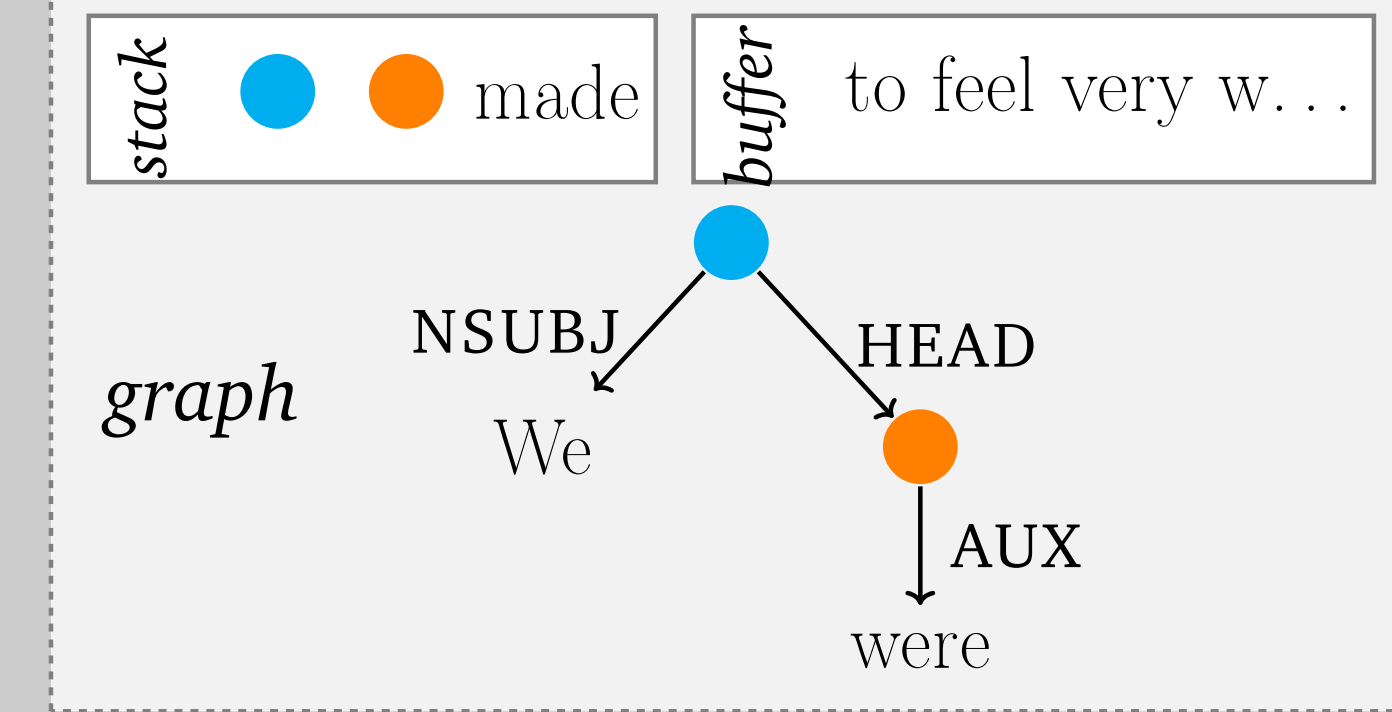
github.com/CoNLL-UD-2018/HUJI

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.

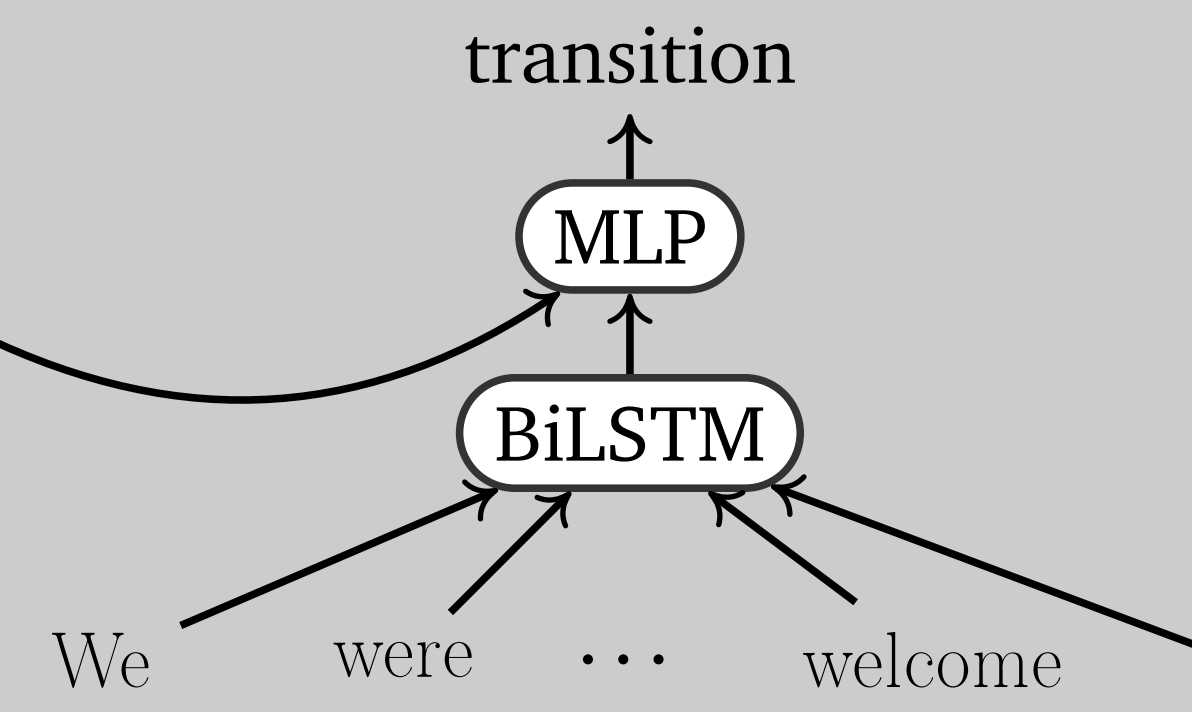
Transitions:

SHIFT, REDUCE, NODE_X, RIGHT-EDGE_X, LEFT-ENHANCED_X, RIGHT-ENHANCED_X, SWAP, FINISH

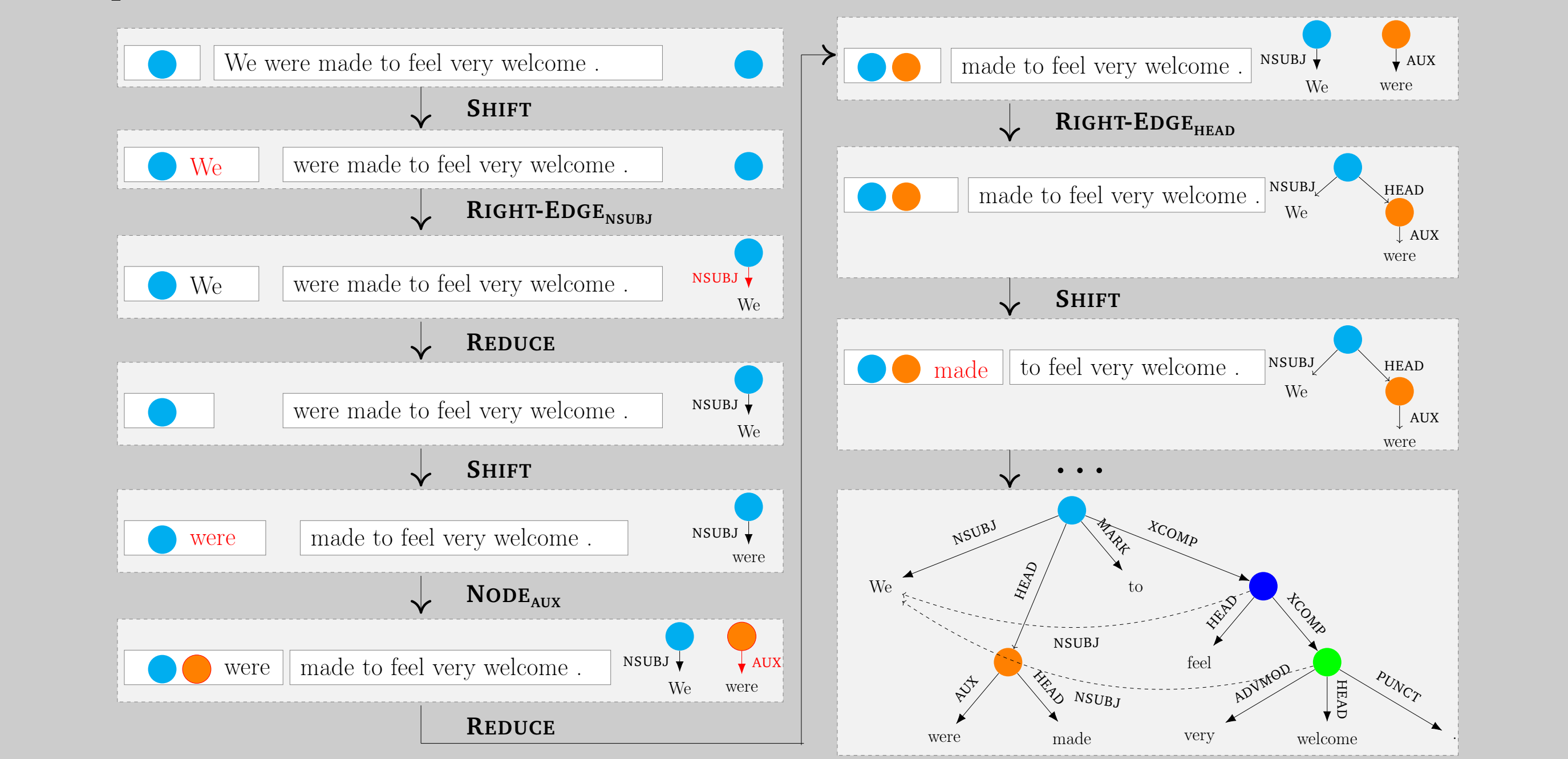
Parser state:



Transition classifier:



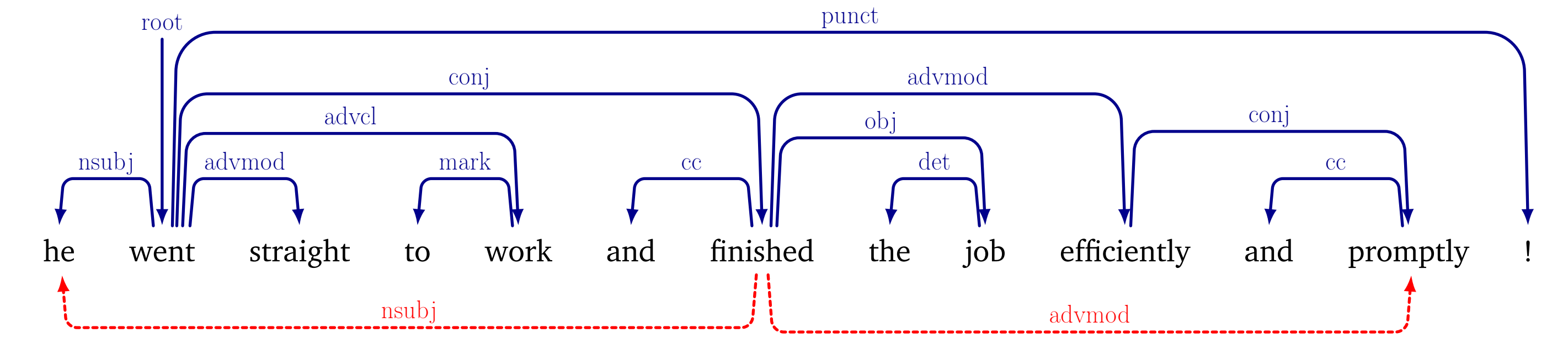
Example:



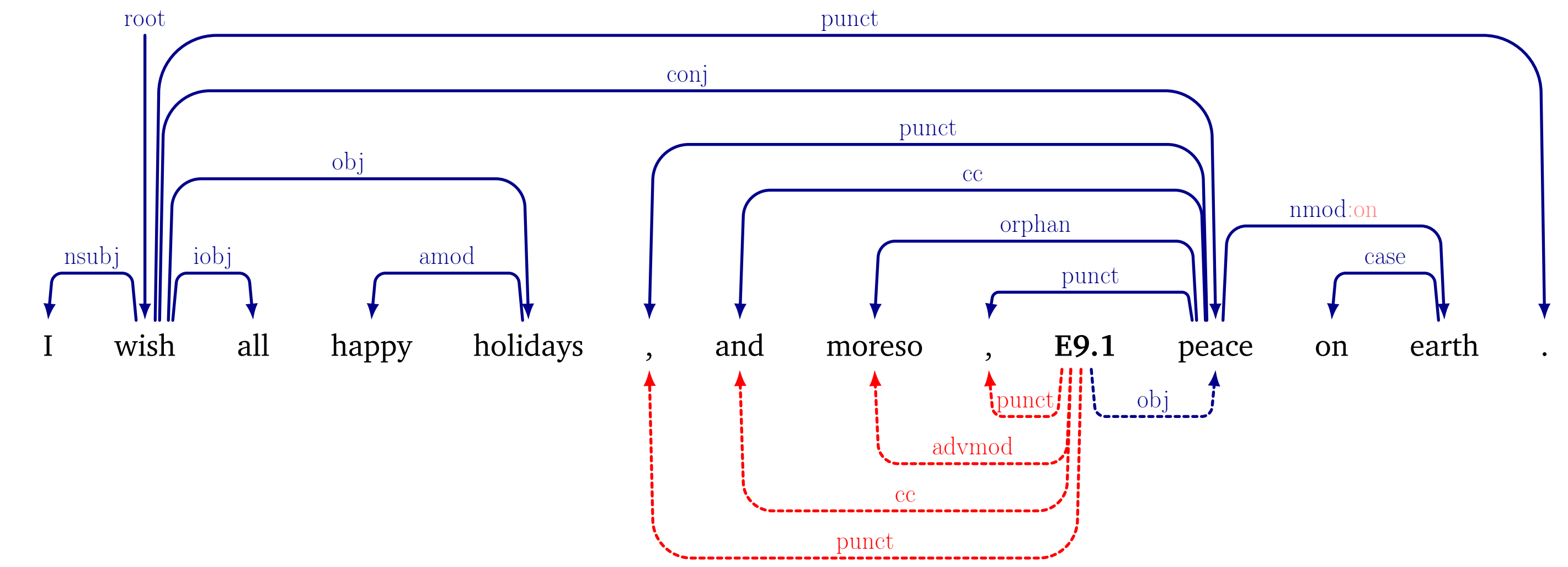
Enhanced Dependencies

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

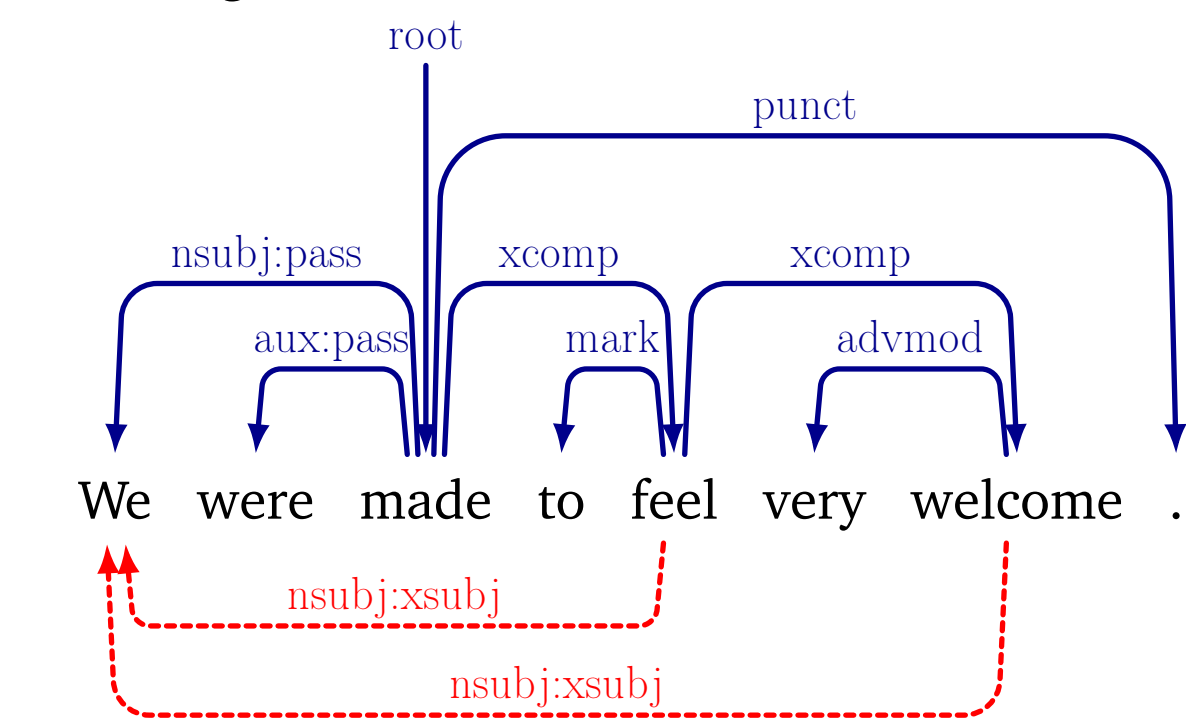
Conjoined predicates and arguments:



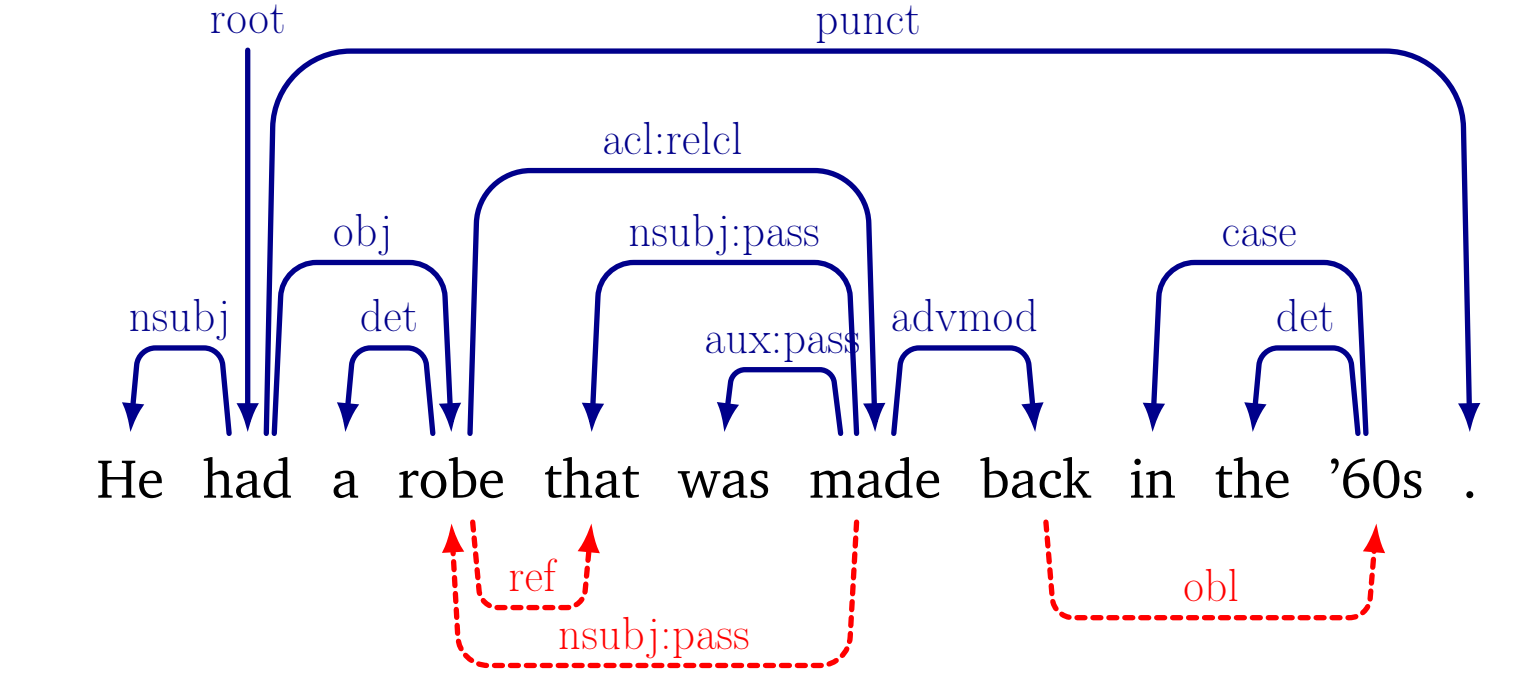
Null nodes due to elided predicates, case information:



Raising:

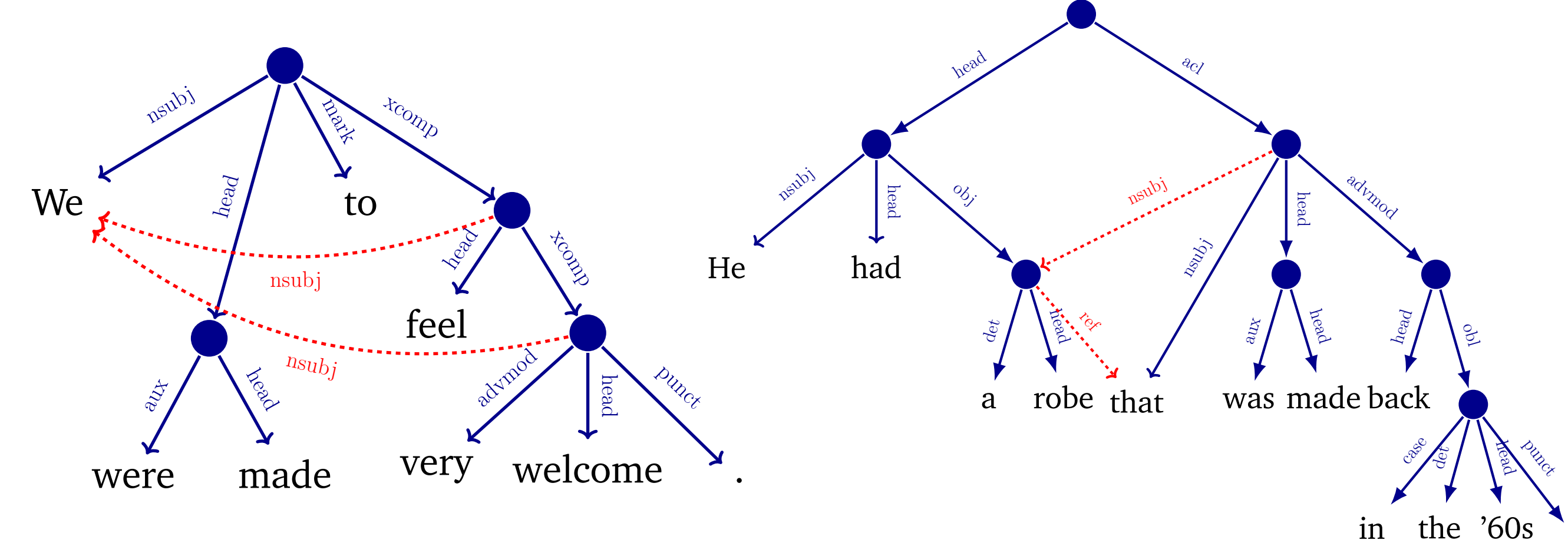


Relative clause:

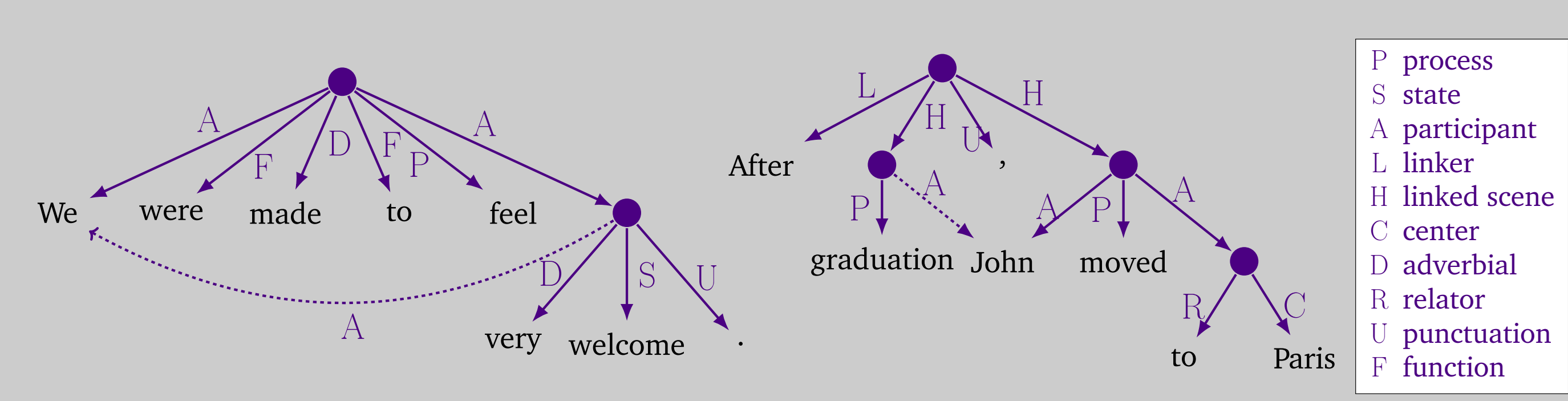


Unified DAG Format

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

	TUPA (official)	TUPA (unofficial)	UDPipe (baseline)	LAS-F1	Enhanced LAS-F1
All treebanks	53.69	58.48	65.80	72.10	57.13
Big treebanks	62.07	67.36	74.14	71.82	54.65
PUD treebanks	56.35	56.82	66.63	69.23	49.12
Small treebanks	36.74	41.19	55.01	72.33	54.54
Low-resource	8.53	12.68	17.17	72.08	0.00
UDPipe				77.62	0.00
UDPipe + CoreNLP				76.66	21.68

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

TUPA: first general parser for enhanced UD.

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

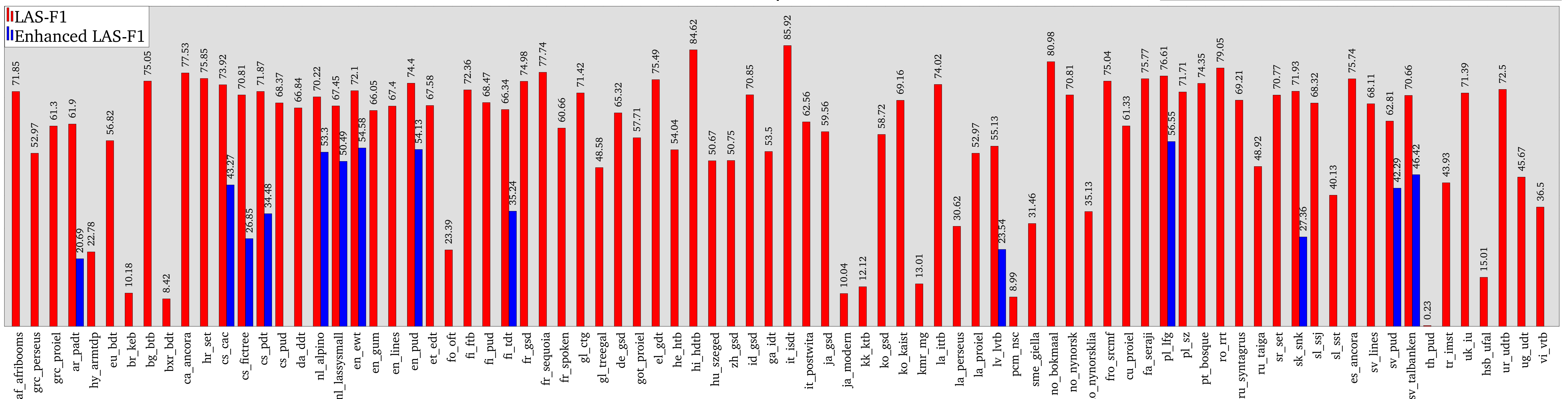
References

- [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.

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



tinyurl.com/semEval-uCCA



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