# Multitask Parsing Across Semantic Representations


**Daniel Hershcovich**[1,2] & **Omri Abend**[2] & **Ari Rappoport**[2]

[1]The Edmond and Lily Safra Center for Brain Sciences
[2]School of Computer Science and Engineering
The Hebrew University of Jerusalem {danielh,oabend,arir}@cs.huji.ac.il


Multitask learning improves UCCA parsing, using AMR, SDP and UD as auxiliary tasks with a general transition-based parser.

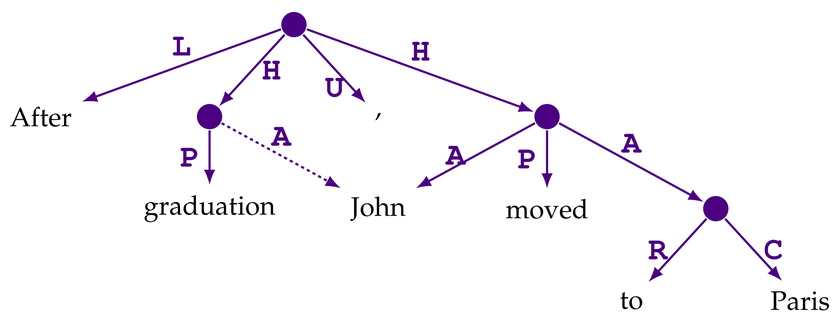
Training data for parsing semantic representations is scarce. We consider four schemes:

**UCCA**: Intuitive, cross-lingual, and modular semantic representation. *Primary edges* form a tree. *Remote edges* (dashed) allow reentrancy, creating a directed acyclic graph [1].

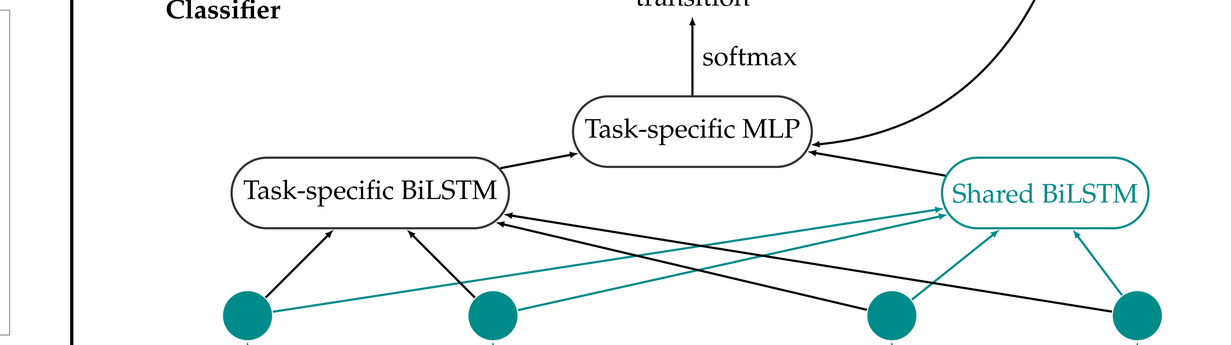
**AMR**: Abstract graph on concepts and constants. Rooted DAG with labeled nodes and edges. Encodes named entities, argument structure, semantic roles, word sense, coreference [3].

**SDP**: Set of related bilexical semantic DAG schemes: DM, PAS, PSD and CCD. We use **DM** (DELPH-IN MRS). Encodes argument structure for many predicate types [7].

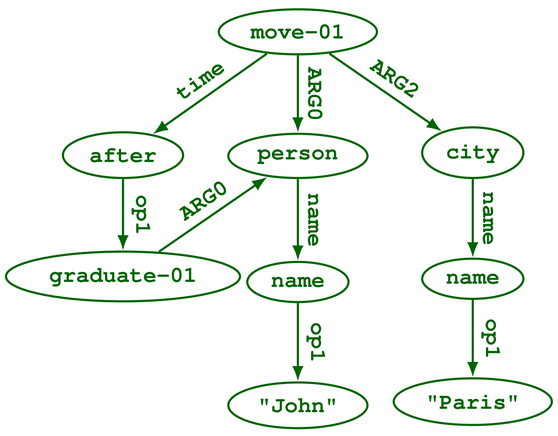
**UD**: Cross-lingual syntactic bilexical tree. Encodes syntactic relations between words [6].
**UD**$^{++}$ (Enhanced++ UD) adds and augments edges, creating a bilexical DAG [8].



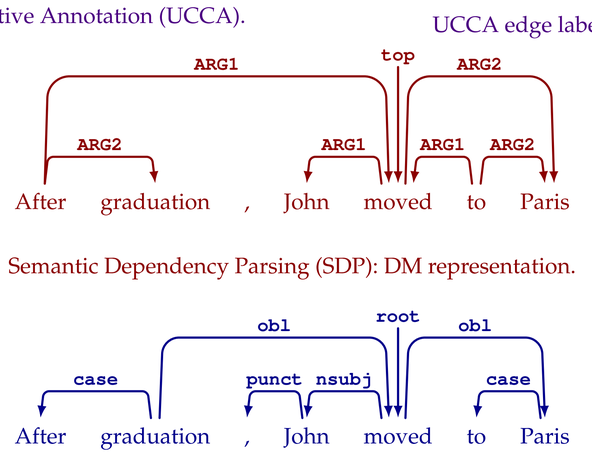

Universal Conceptual Cognitive Annotation (UCCA).




UCCA edge labels.




Abstract Meaning Representation (AMR).




Semantic Dependency Parsing (SDP): DM representation.

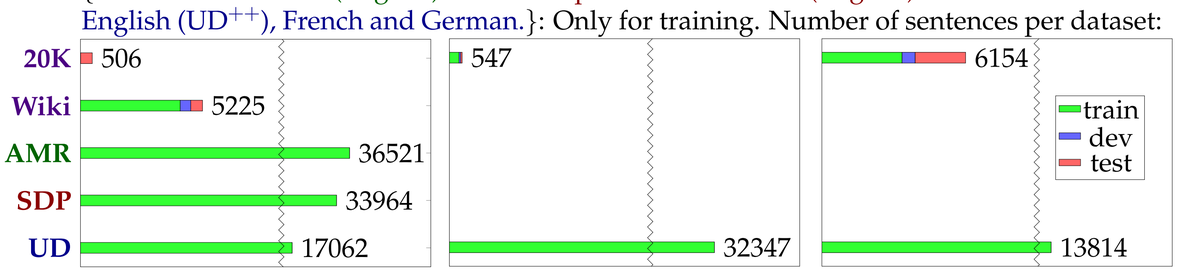



Universal Dependencies (UD).

Semantic representations share much of their content [2].
**Multitask learning** exploits task overlap, effectively extending the training data.
We focus on UCCA parsing due to its small training set.
As auxiliary tasks, we use **unlabeled** AMR, SDP and UD parsing.

## Data

**UCCA**: (1) English Wikipedia (**Wiki**); (2) *Twenty Thousand Leagues Under the Sea* (**20K**), annotated in English (small, only test) French (small), and German (pre-release, noisy). {**AMR**: LDC2017T10 (English). **SDP**: DM part from SDP 2016 (English). **UD**: v2.1 treebanks: English (UD$^{++}$), French and German.}: Only for training. Number of sentences per dataset:



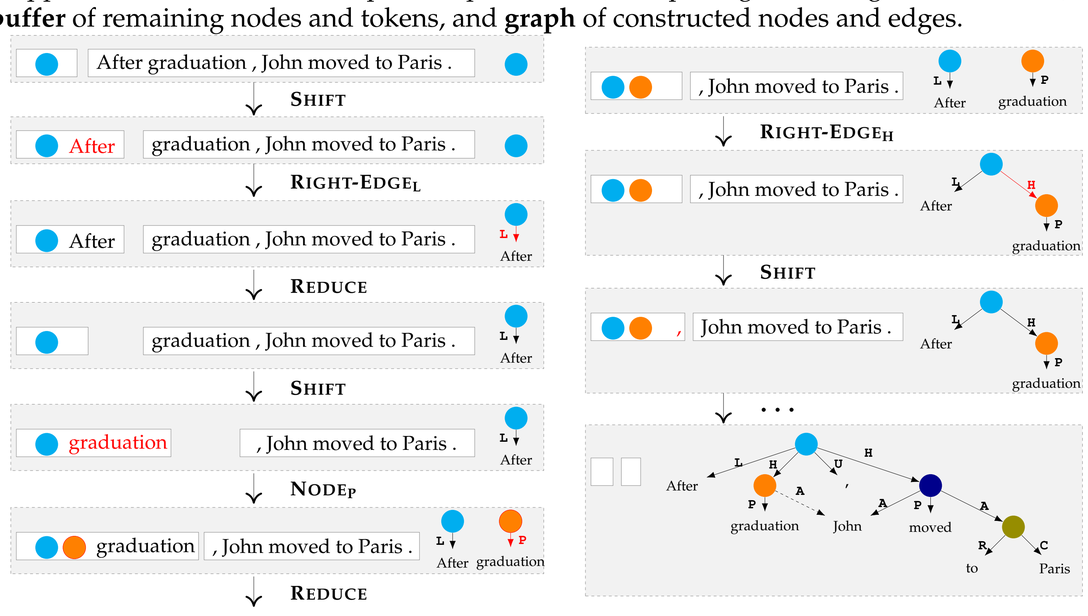

**English** corpora. **French** corpora. **German** corpora.

Domains differ, too:

| UCCA | AMR | SDP | UD |
|---|---|---|---|
| Wikipedia | blogs | news | blogs |
| books | news | | news |
| | emails | | emails |
| | reviews | | reviews |
| | | | Q&A |

## TUPA: A Transition-Based DAG Parser

We extend a UCCA parser supporting reentrancy, discontinuity and non-terminal nodes [4].
It applies a *transition* at each step to the parser state, comprising a working **stack** of nodes, **buffer** of remaining nodes and tokens, and **graph** of constructed nodes and edges.




## Transition Classifier

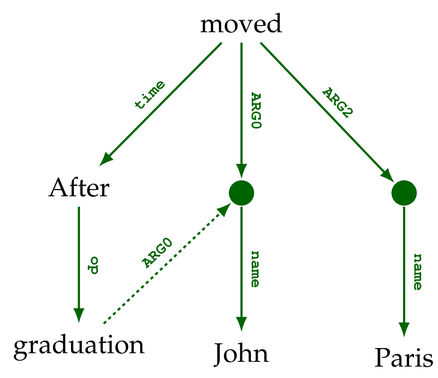
Bidirectional LSTM RNN to encode text token features + feedforward NN for classification.
Multitask architecture: **Task-specific** BiLSTM for the main task + **shared** BiLSTM across all tasks. Concatenated to select each transition using a task-specific feedforward NN.



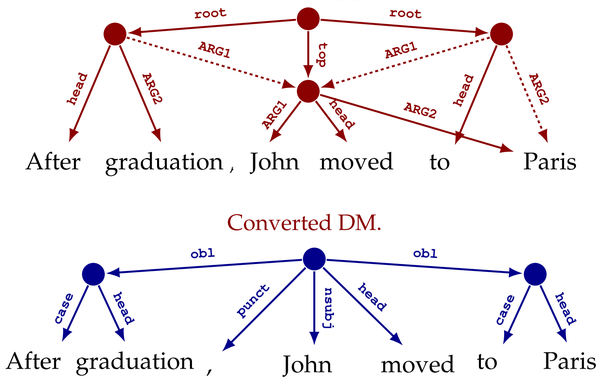

Limited capacity promotes generalization by using the shared parameters for all tasks [5].

## Unified DAG Format

We convert all representations into a format similar to UCCA and supported by TUPA.



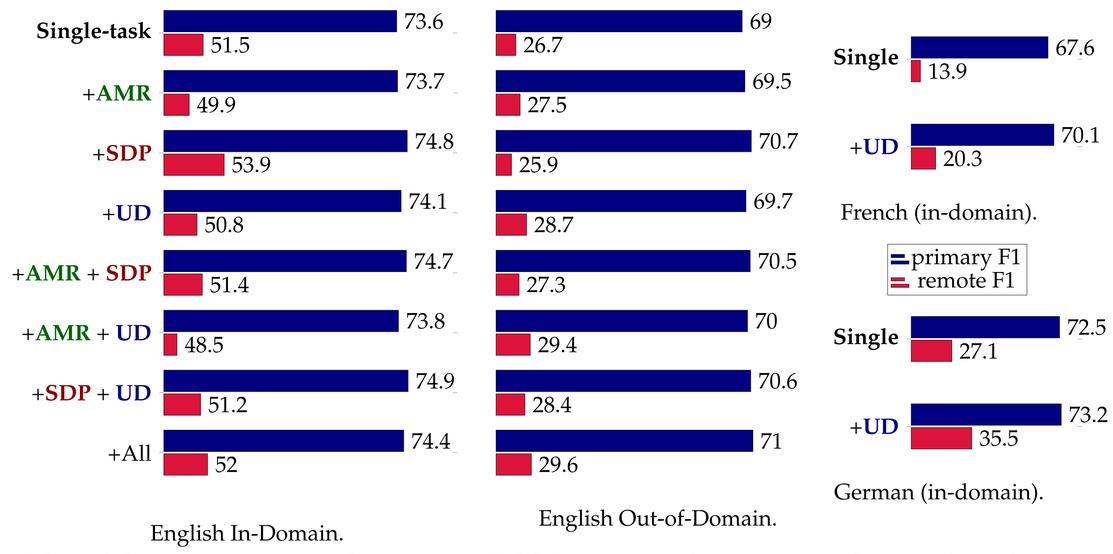

Converted AMR.




Converted UD.

## Experiments

**English**. Train: UCCA Wiki (+aux), test: UCCA Wiki (in-domain) or 20K (out-of-domain).
**French** and **German**. Train: 20K (+UD as aux), test: 20K (both in-domain).

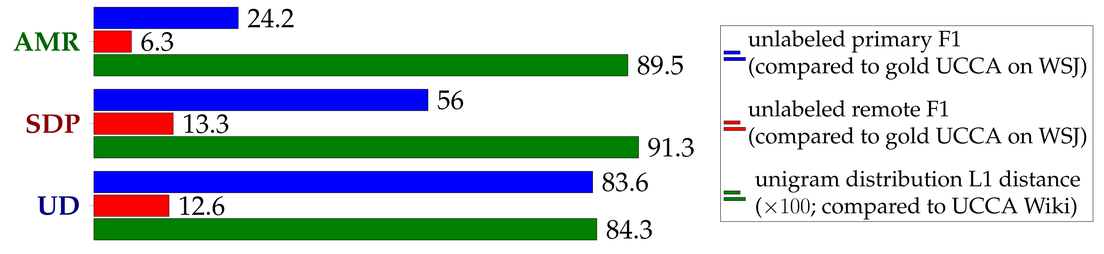



English In-Domain. English Out-of-Domain. French (in-domain). German (in-domain).

Multitask learning consistently improve UCCA parsing when compared to single-task.

## Task Similarity

Does improvement depend on structural task similarity, or training corpus similarity?
We compared **annotations of 100 WSJ sentences**, and **training corpus word distributions**.





### References

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

[2] Omri Abend and Ari Rappoport. The state of the art in semantic representation. In *Proc. of ACL*, pages 77–89, 2017.

[3] Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Martha Palmer, and Nathan Schneider. Abstract Meaning Representation for sembanking. In *Proc. of the Linguistic Annotation Workshop*, 2013.

[4] Daniel Hershcovich, Omri Abend, and Ari Rappoport. A transition-based directed acyclic graph parser for UCCA. In *Proc. of ACL*, pages 1127–1138, 2017.

[5] Héctor Martínez Alonso and Barbara Plank. When is multitask learning effective? Semantic sequence prediction under varying data conditions. In *Proc. of EACL*, pages 44–53, 2017.

[6] Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. Universal dependencies v1: A multilingual treebank collection. In *Proc. of LREC*, may 2016.

[7] Stephan Oepen, Marco Kuhlmann, Yusuke Miyao, Daniel Zeman, Silvie Cinkova, Dan Flickinger, Jan Hajic, Angelina Ivanova, and Zdenka Uresova. Towards comparability of linguistic graph banks for semantic parsing. In *Proc. of LREC*, may 2016.

[8] 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