Frustratingly Simple but Surprisingly Strong: Using Language-Independent Features for Zero-shot Cross-lingual Semantic Parsing

Jingfeng Yang, Federico Fancellu, Bonnie Webber, Diyi Yang

Georgia Tech College of Computing
School of Interactive Computing

The University of Edinburgh

(DRS) Semantic Parsing

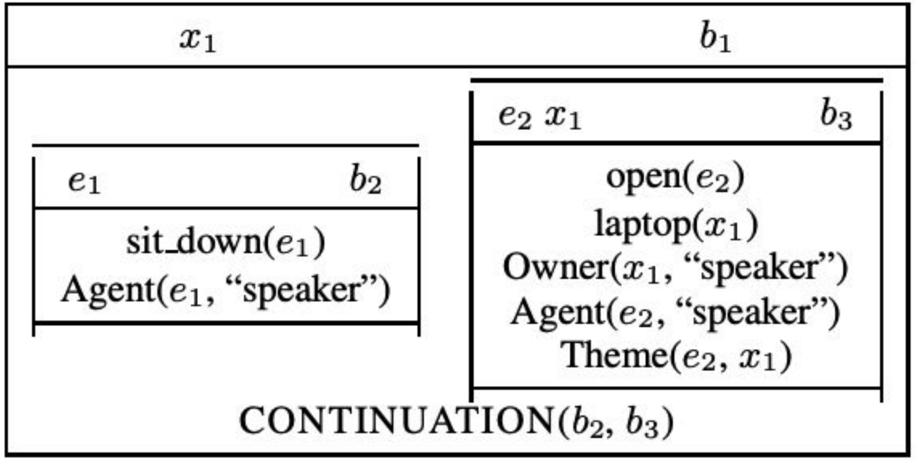


Figure 1: The Discourse Representation Structure (DRS) for "*I sat down and opened my laptop*". For simplicity, we have omitted any time reference.

Problem and Question

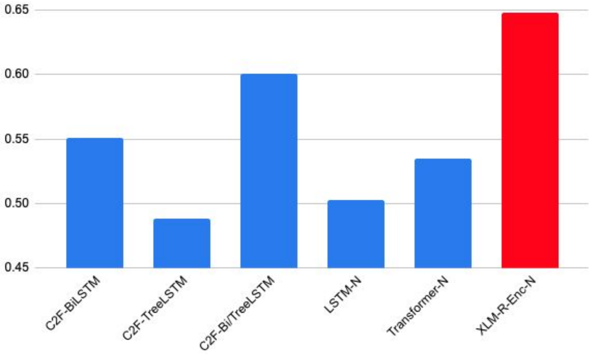
- Annotating meaning representations is expensive, often requiring experts' (linguistists) domain knowledge.
- Limited meaning representation data annotated in languages other than English.
- What does it require to train a semantic parser on English annotated data, and test (use) it on other languages? (Zero-shot Cross-lingual Semantic Parsing)
- Pretrained Large-scale Multilingual models (mBERT, XLM, XLM-R etc.)?
- What else?

Prior Work on Cross-lingual Semantic Parsing

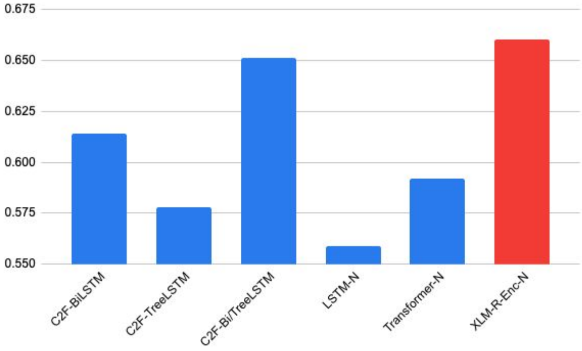
- Parallel Corpus, Machine Translation, Projection
  Padó and Lapata, 2005; Damonte and Cohen, 2017; Zhang et al., 2018; Xu et al., 2020

- Parameter-Shared Neural Models
  - Cross-lingual word embeddings
    Zhu et al., 2020; Li et al., 2020; Oepen et al., 2020
  - Large-scale pretrained multilingual models
    Duong et al., 2017; Susanto and Lu, 2017; Mulcaire et al., 2018; Hershcovich et al., 2019; Cai and Lapata, 2020

6 DRS Semantic Parsers tested on 3 Languages (trained on English)



German (PMB 2.1)

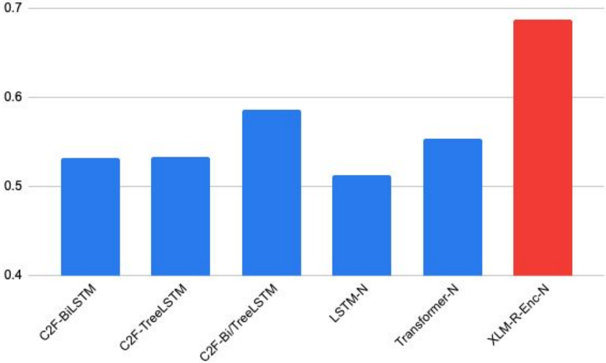


Italian (PMB 2.1)

Sequential Encoder / Tree Encoder (TreeLSTM)

Sequential Decoder / Coarse2fine Decoder (C2F)

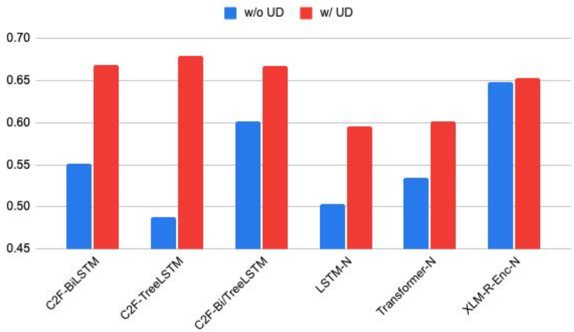
Cross-lingual Word Embeddings /
Large-scale Pretrained Multilingual Model (XLM-R)



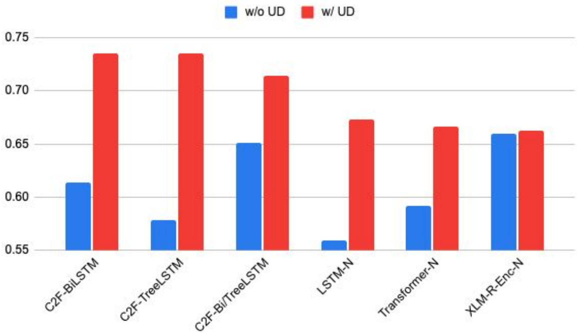
Dutch (PMB 2.1)

Universal Dependency as Language-Independent Features

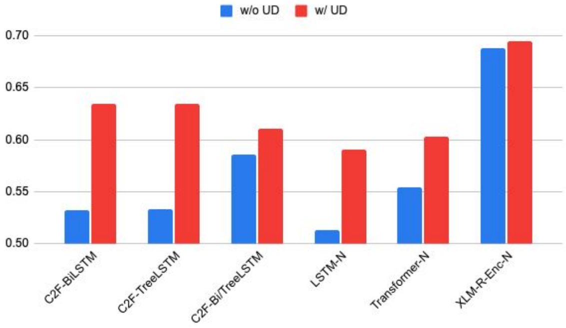
Universal Dependency helps a lot !



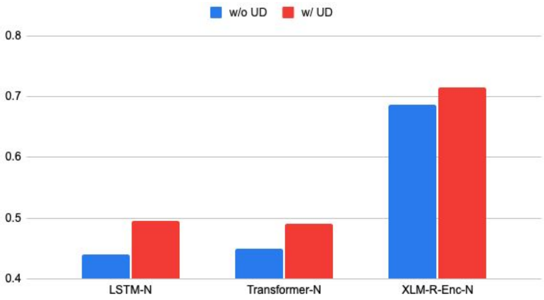
German (PMB 2.1)



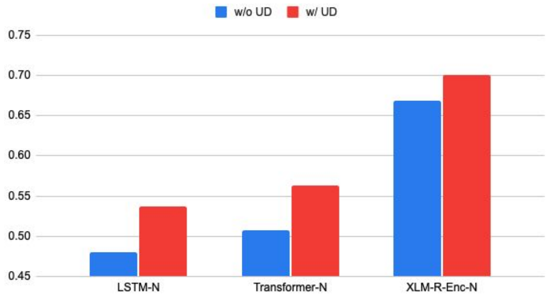
Italian (PMB 2.1)



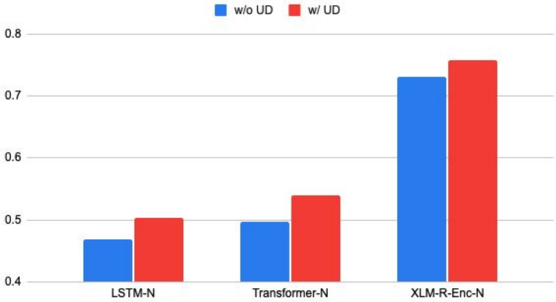
Dutch (PMB 2.1)



German (PMB 3.0)



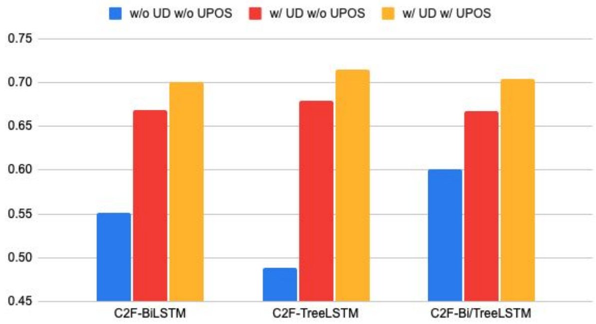
Italian (PMB 3.0)



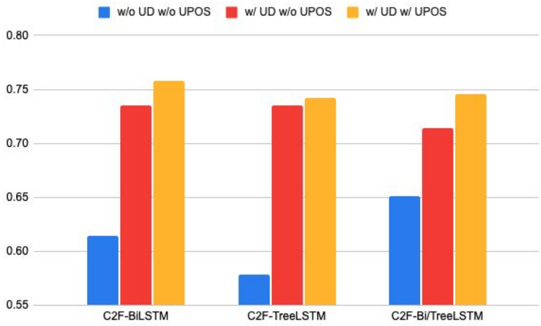
Dutch (PMB 3.0)

Universal POS Tags as Language-Independent Features

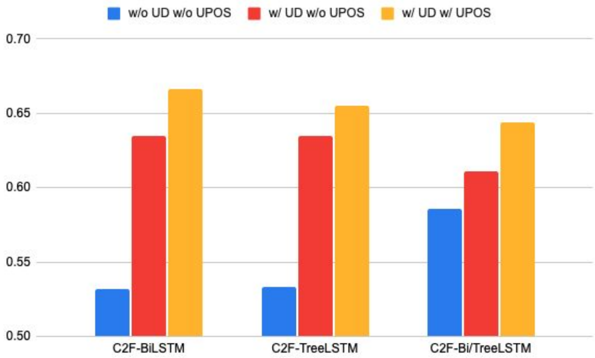
Universal POS tags also help.



German (PMB 2.1)



Italian (PMB 2.1)



Dutch (PMB 2.1)

Model Efficiency Comparison

Simple coarse2fine decoding parsers with UD have competitive performance and much higher efficiency compared with the large pretrained multilingual model.

	TT/min	PT/s	Size/P
XLM-R-Enc-N	300	156	773.3M
C2F-BiLSTM$_D$	9(33x)	25(6.2x)	2.5M(309x)
C2F-BiLSTM$_{W,D}$	10(30x)	27(5.8x)	168.7M(4.6x)

Table 2: Model size and running time in PMB 2.1, where TT is training time till convergence, PT is Parsing/Inference time in German test set and Size is the model size measured with number of Parameters (P).

Why Does Universal Dependency Help a Lot?

Gold DRS: DRS(schläfrig(S1) TIME(S1 T1) THEME(S1 Z0) Speaker(Z0) Time(T1) EQ(T1 Y0) Now(Y0))

DRS before adding UD: DRS(ich(X1) bin(X2) TIME(E1 T1) THEME(E1 X1) Time(T1) EQ(T1 Y0) Now(Y0))

DRS after adding UD: DRS(Person(X1) schläfrig(S1) TIME(S1 T1) THEME(S1 X1) Time(T1) EQ(T1 Y0) Now(Y0))

Figure 1: The Discourse Representation Structure (DRS) for "*Ich bin schläfrig. (I am sleepy.)*", including ground truth DRS, parsed DRS before and after adding UD relations as features.

Although a parser could understand the coarse meaning of the whole sentence without UD, it struggles with understanding some lexical-level meaning.

However, it successfully identifies "schläfrig" as the event, based on its UD relation tag "root".

Thank you!

Project Page:

Project Link:
https://github.com/GT-SALT/Multilingual-DRS-Semantic-Parsing