

A Appendices for “Cross-Lingual BERT Transformation for Zero-Shot Dependency Parsing”

A.1 Statistics of UD (v2.2) Treebanks

The statistics of the Universal Dependency treebanks we used are summarized in Table 1.

Language	Language Family	Treebank	Test Sentences
English (en)	IE.Germanic	EWT	2,077
German (de)	IE.Germanic	GSD	977
Danish (da)	IE.Germanic	DDT	565
Swedish (sv)	IE.Germanic	Talbanken	1,219
Dutch (nl)	IE.Germanic	Alpino, LassySmall	1,472
French (fr)	IE.Romance	GSD	416
Italian (it)	IE.Romance	ISDT	482
Spanish (es)	IE.Romance	GSD, AnCora	2,147
Portuguese (pt)	IE.Romance	Bosque, GSD	1,681
Romanian (ro)	IE.Romance	RRT	729
Slovak (sk)	IE.Slavic	SNK	1,061
Polish (pl)	IE.Slavic	LFG, SZ	2,827
Bulgarian (bg)	IE.Slavic	BTB	1,116
Slovenian (sl)	IE.Slavic	SSJ, SST	1,898
Czech (cs)	IE.Slavic	PDT, CAC, CLTT, FicTree	12,203
Finnish (fi)	Uralic	TDT	1,555
Estonian (et)	Uralic	EDT	2,737
Latvian (lv)	IE.Baltic	LVTB	1,228

Table 1: Statistics of the Universal Dependency treebanks we selected in our experiments. For language family, IE is the abbreviation for Indo-European.

A.2 Implementation Details

For the graph-based Biaffine parser, we exclude the learned embeddings in our re-implementation, to focus on the effect of pre-trained embeddings. Besides, the universal POS tags are used throughout our experiments.

The PyTorch version of the base BERT model for English and multi-languages¹ are used to generate the 768-dimensional contextualized embeddings for English and target languages respectively. In the GD-based method, we use Adam optimizer, with a learning rate of 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$.

A.3 Full Results on UD Treebanks

The LAS of our models (including the combination of cross-lingual FastText embeddings and our CLBT ones, where they are concatenated as the input to the parser) and the baseline ones are shown in Table 2, and UAS in Table 3.

¹github.com/huggingface/pytorch-pretrained-BERT

Lan.	Static	Contextualized				
	FT-SVD	mBERT	CLBT (SVD)	CLBT (SVD) +FT	CLBT (GD)	CLBT (GD) +FT
en	88.31	90.71	91.03	91.32	91.03	91.32
de	59.31	63.41	64.47	64.78	62.14	63.05
da	68.81	70.57	71.60	72.03	71.66	71.57
sv	73.49	70.09	73.33	75.70	75.95	76.72
nl	60.11	65.66	65.45	65.90	63.86	64.92
fr	73.46	72.97	74.70	75.56	76.59	76.38
it	76.23	79.02	79.46	79.18	78.98	79.27
es	66.91	65.43	67.14	67.47	68.33	67.71
pt	67.98	67.11	69.12	69.00	69.25	69.09
ro	52.11	46.40	55.14	54.79	55.84	55.53
sk	56.98	50.76	59.46	59.43	59.92	59.60
pl	58.59	63.10	65.37	65.71	65.80	66.80
bg	66.68	71.20	70.26	70.33	70.75	70.89
sl	54.57	56.78	57.42	57.36	57.21	57.68
cs	52.80	45.20	52.20	52.37	52.99	53.05
fi	48.74	49.56	51.00	53.26	52.61	53.91
et	44.40	46.64	47.79	48.27	48.52	48.57
lv	49.59	45.11	48.59	50.04	49.78	50.98
AVG.	60.63	60.53	63.09	63.60	63.54	63.87

Table 2: Results (LAS%) on the test sets. The two columns on the left show results of baseline models, while the others on the right show results of our models. Languages are split by language families with dashed lines. AVG. means the average of results from all target languages. (**Lan.** stands for Language, **FT** stands for FastText.)

Lan.	Static	Contextualized				
	FT-SVD	mBERT	CLBT (SVD)	CLBT (SVD) +FT	CLBT (GD)	CLBT (GD) +FT
en	90.44	92.49	92.81	93.11	92.81	93.11
de	69.49	72.34	73.72	73.72	71.08	71.51
da	77.36	79.29	79.63	80.05	79.16	79.70
sv	81.23	78.25	80.57	82.28	82.64	83.34
nl	67.88	73.22	72.80	73.30	71.00	72.11
fr	78.35	78.79	80.01	81.10	80.85	80.92
it	81.10	83.73	84.53	84.22	83.33	83.95
es	74.92	73.97	75.52	75.89	75.70	75.59
pt	76.46	75.09	77.17	76.90	76.71	76.44
ro	63.23	58.45	66.01	66.07	66.30	66.00
sk	65.41	60.19	67.56	68.31	67.62	67.83
pl	71.89	74.03	76.68	76.25	76.52	77.04
bg	78.05	82.83	82.14	82.01	81.51	81.70
sl	66.27	67.86	69.04	69.16	68.26	68.59
cs	61.88	54.86	61.02	61.29	61.26	61.26
fi	66.36	65.45	65.65	68.28	67.96	69.16
et	65.25	64.22	65.26	65.87	66.76	66.49
lv	71.43	61.73	65.54	66.98	67.41	68.20
AVG.	71.56	70.84	73.11	73.63	73.18	73.52

Table 3: Results (UAS%) on the test sets. The two columns on the left show results of baseline models, while the others on the right show results of our models. AVG. means the average of results from all target languages. (**Lan.** stands for Language, **FT** stands for FastText.)