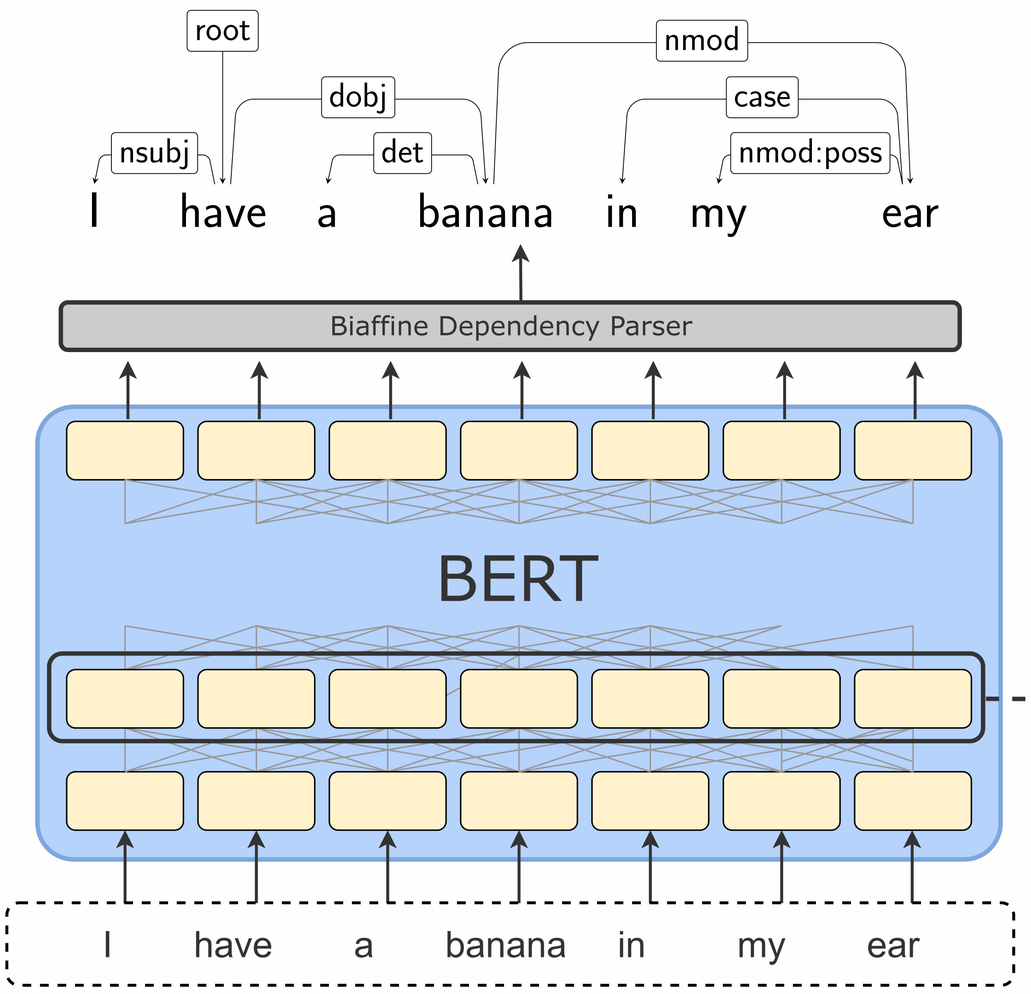
Language-Specific Adaptations of Multilingual Language Models for Universal Dependency Parsing

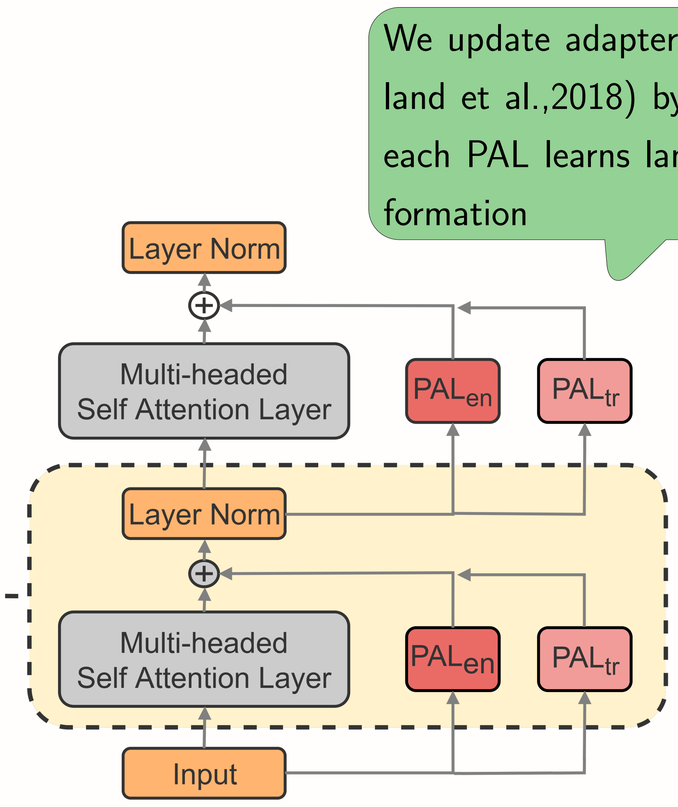
rijksuniversiteit groningen

Ahmet Üstün, Arianna Bisazza, Gosse Bouma, Gertjan van Noord

NWO



We extended Udify (Kondratyuk et al.,2019) that uses pretrained multilingual BERT as sentence encoder in biaffine parser



Overview

Task: Adapting mBERT-parser to N different languages

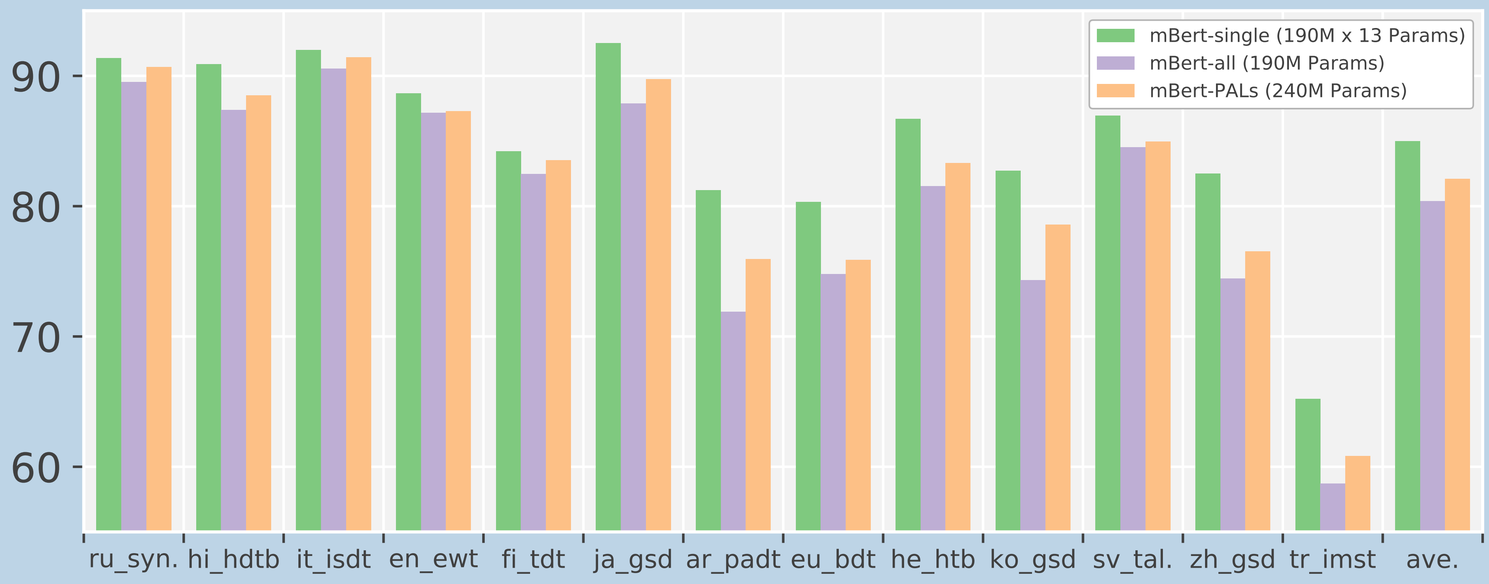
Problem: One parser trained on all languages < monolingual parsers trained on each language

Goal: Improve the multilingual parser by capturing language specific information in specific modules (PALs)

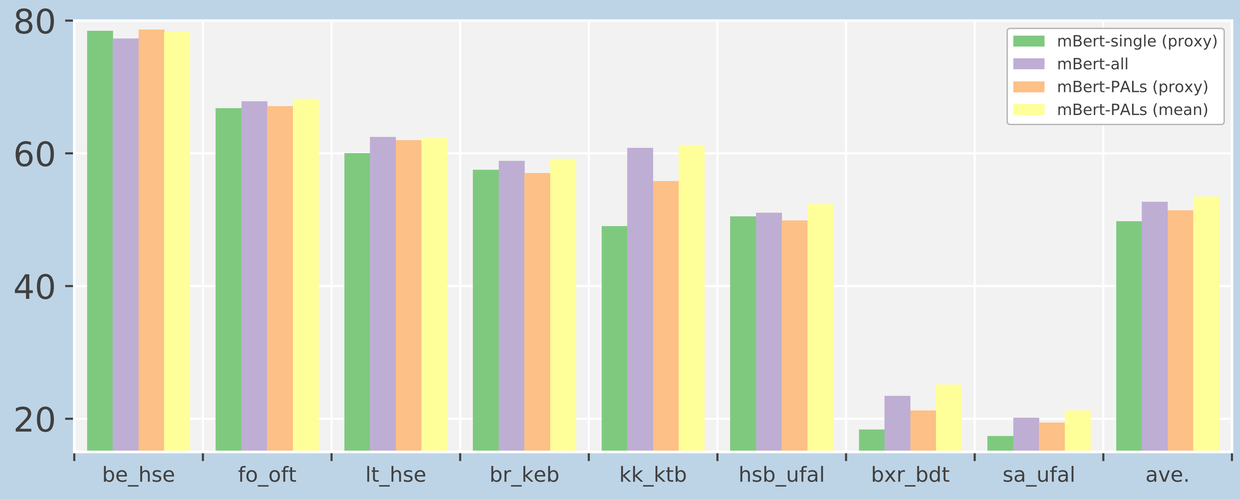
Languages with Training Data

Dataset	Family	Size(K)
ru_syn.	IE, Slavic	48.8
hi_hdtb	IE, Indic	13.3
it_isdt	IE, Romance	13.1
en_ewt	IE, Germanic	12.5
fi_tdt	Uralic, Finnic	12.5
ja_gsd	Japanese	7.1
ar_padt	Afro-Asiatic, Semitic	6.1
eu_bdt	Basque	5.4
he_htb	Afro-Asiatic, Semitic	5.2
ko_gsd	Korean	4.4
sv_tal.	IE, Germanic	4.3
zh_gds	Sino-Tibetan	4.0
tr_imst	Turkic, Southwestern	3.7

We train the parser either separately (`mBert-single`) or on all treebanks (`mBert-all`) with/out adapters (`mBert-PALs`).



Languages without Training Data (Zero-Shot)



We use `mBert-single`/`mBert-PALs` with a proxy language or we take the mean of each PALs' activations for all languages.

Language	Belarusian	Faroese	Lithuanian	Breton	Kazakh	U. Sorbian	Buryat	Sanskrit
Dataset	be_hse	fo_oft	lt_hse	br_keb	kk_ktb	hsb_ufal	bxr_bdt	sa_ufal
Proxy	ru_syn.	en_ewt	ru_syn.	en_ewt	tr_imst	ru_syn.	ru_syn.	hi_hdtb

Conclusion

- Language adapters (PALs) help the model to learn language-specific information that increase the parsing performance when model is trained on various languages all together.
- Training model with PALs on all languages still underperforms compared to the single language training.
- Without training data (zero-shot), using proxy languages (`mBert-single`/`mBert-PALs`) does not increase performance.
- Using adapters of all languages, `mBert-PALs(mean)`, increases parsing scores which indicates the future direction of our research.