Cross-lingual Dependency Parsing with Unlabeled Auxiliary Languages

Wasi Ahmad, Zhisong Zhang, Xuezhe Ma,
Kai-Wei Chang, and Nanyun Peng.

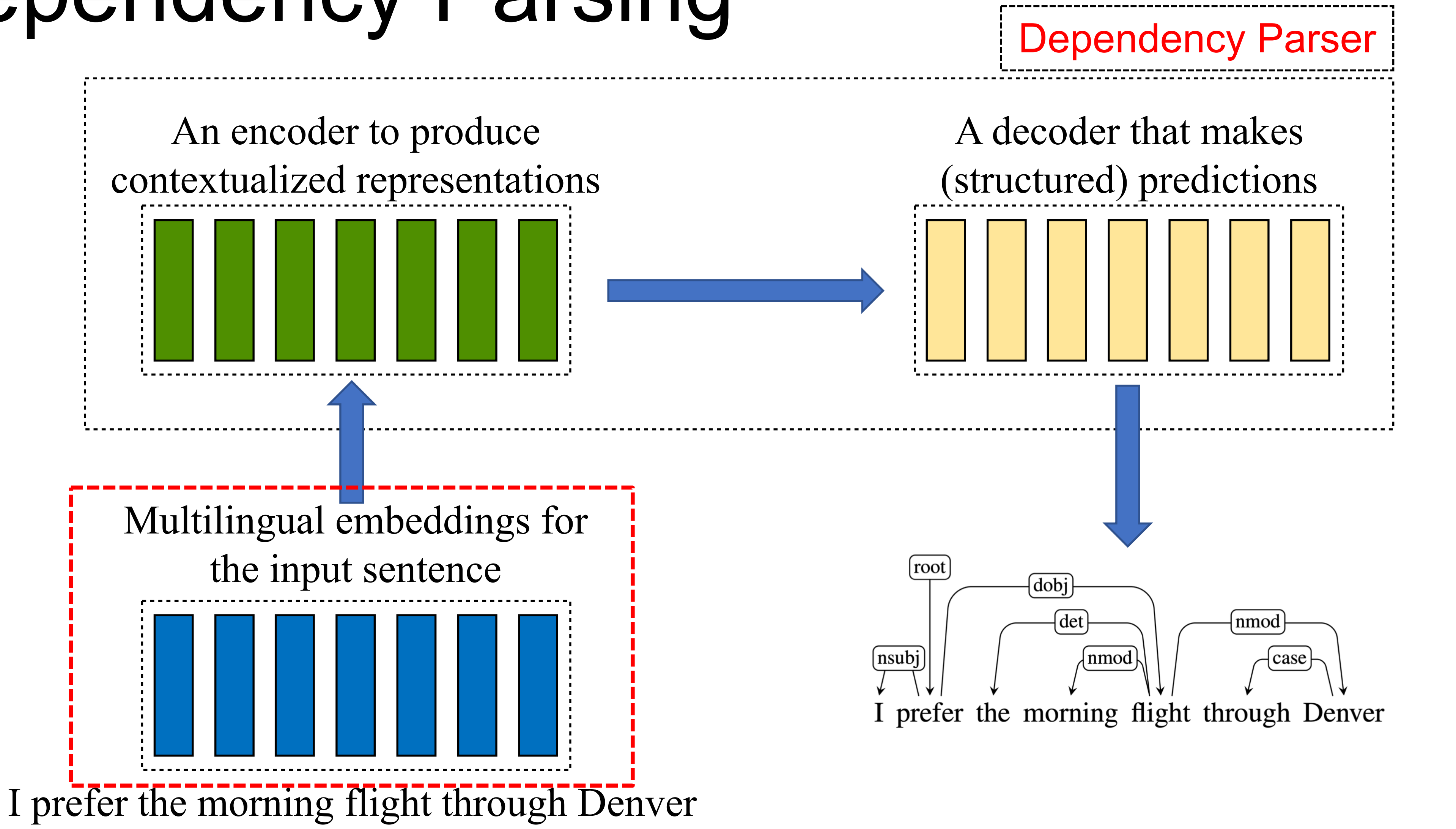
CoNLL, 2019

Motivation

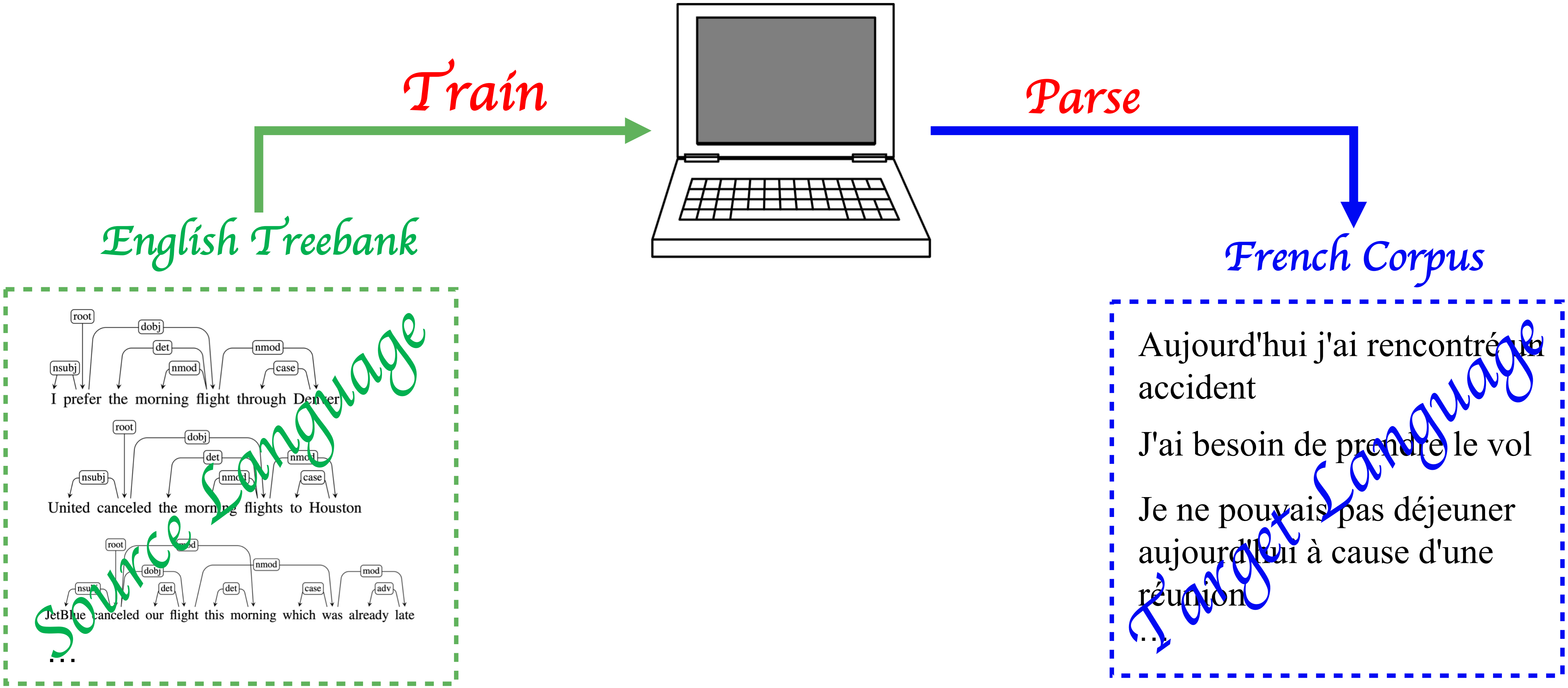
Different languages have different properties
(e.g., word order)

Improve transfer learning across languages
(Learning language-agnostic representation)

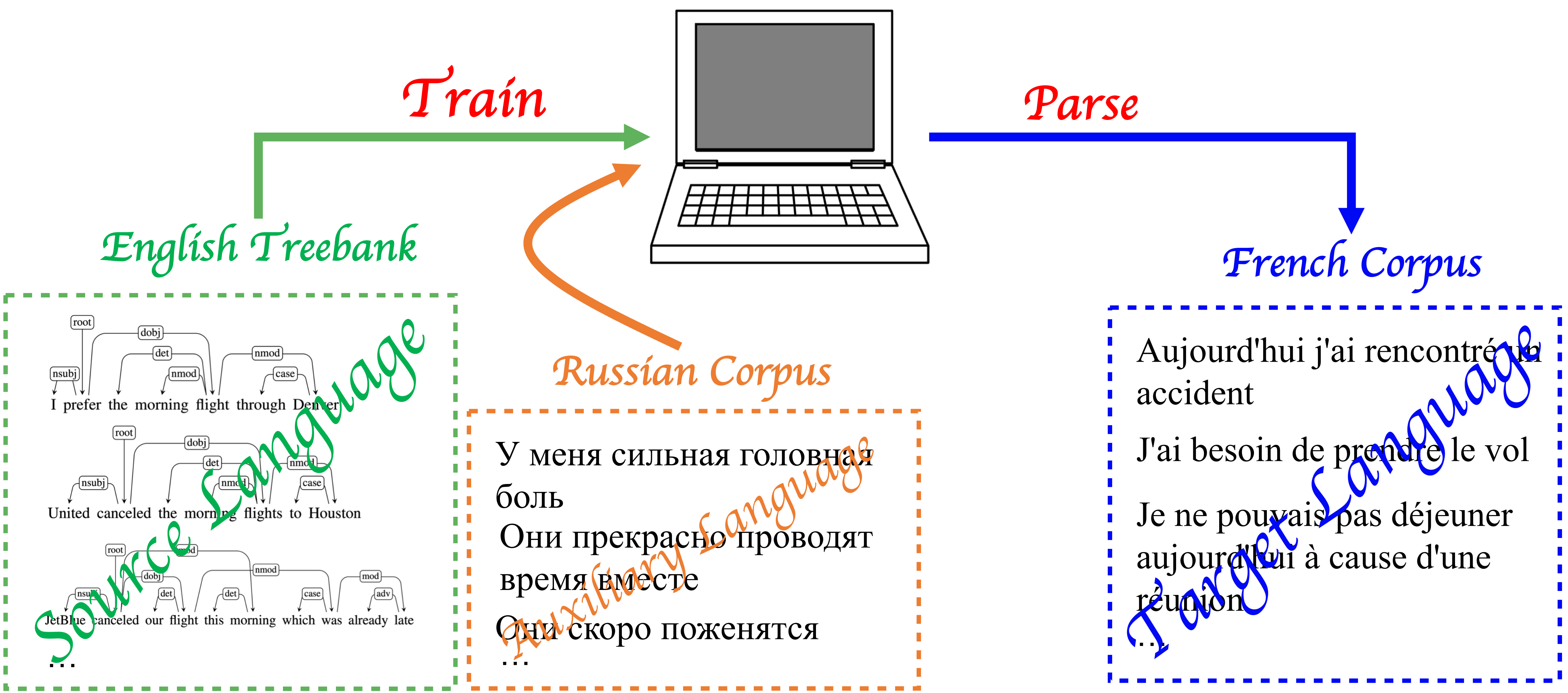
Dependency Parsing



Cross-lingual Dependency Parsing

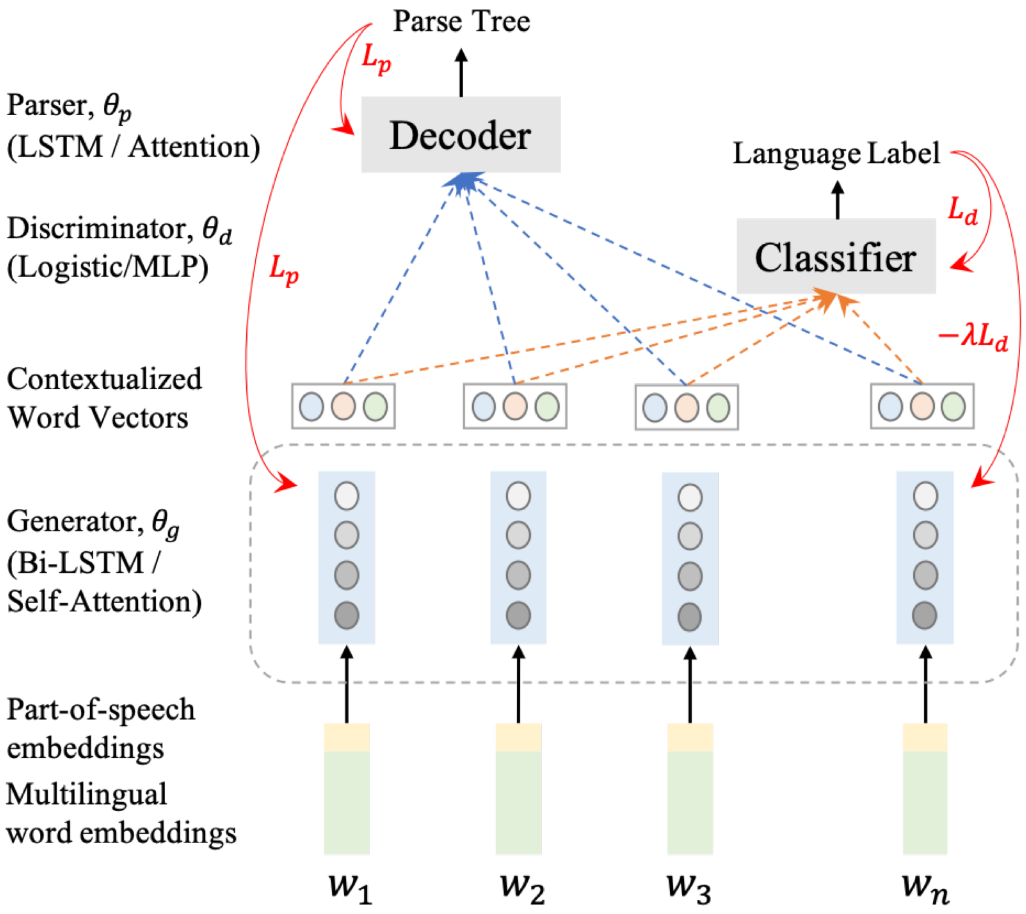


Main Idea


Train
Parse
English Treebank
Source Language
I prefer the morning flight through Denver
United canceled the morning flights to Houston
JetBlue canceled our flight this morning which was already late
…
Russian Corpus
Auxiliary Language
У меня сильная головная боль
Они прекрасно проводят время вместе
Они скоро поженятся
…
French Corpus
Target Language
Aujourd'hui j'ai rencontré un accident
J'ai besoin de prendre le vol
Je ne pouvais pas déjeuner aujourd'hui à cause d'une réunion
…

Main Idea

- Use unlabeled corpora of auxiliary languages
- Adversarial training to learn language-agnostic representation
  - <u>Discriminator</u>: predicts language label



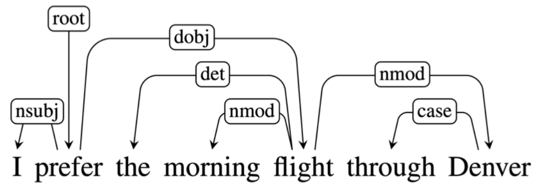
Training Procedure

Step1. Warm-start the parser

- Mini-batch training using source language treebank
- For k iterations

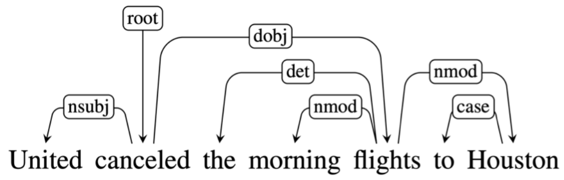
I prefer the morning flight through Denver

Parser



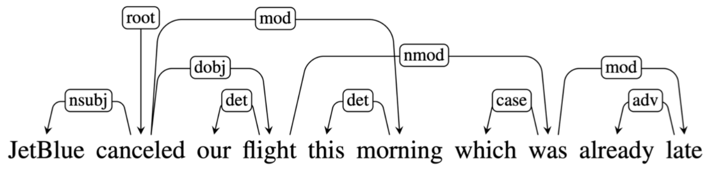
United canceled the morning flights to Houston

Parser



JetBlue canceled our flight this morning which was already late

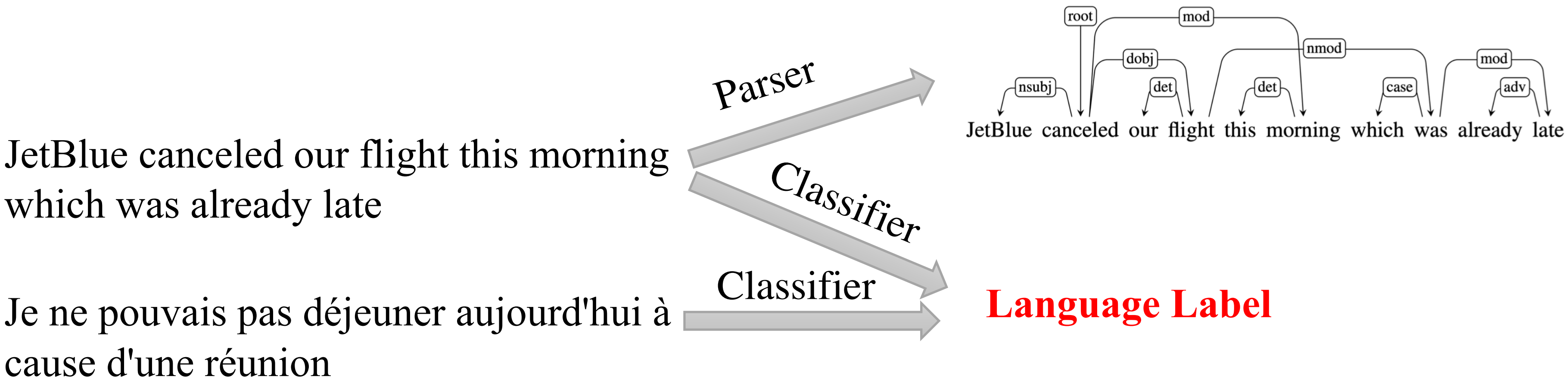
Parser



Training Procedure

Step2. Jointly train using auxiliary languages

- Train the parser on source language
- Adversarial training on both source and auxiliary languages



Experiment Setup

Embedding

- Token embeddings
  - Multilingual Embeddings (MUSE) [Smith et al., 2017, Bojanowski et al., 2017]
  - Multilingual BERT (M-BERT) [Devlin et al., 2017]
- Part-of-speech embeddings

Parsers [Ahmad et al., 2019]

- Graph-based: Self-attentive-Graph
  - Multi-Head Self-Attention (**order-free**)
- Transition-based: RNN-StackPtr
  - BiLSTMs (**order-dependent**)

Experiment Setup

Single Source Transfer Parsing

- Train parser on one language
- Source language: English

Adversarial training

- Using a language pair (one source and one auxiliary language)
- Auxiliary languages are selected based on -
  - Covering different language families
  - Average distance between auxiliary language and all target languages

Experiment Setup

Datasets

- Universal Dependency Treebanks (v2.2)
- 1 Source language, 28 target languages
- 10 families

Evaluation

- Evaluate directly on the target languages (zero-shot)
- Metrics: UAS, LAS

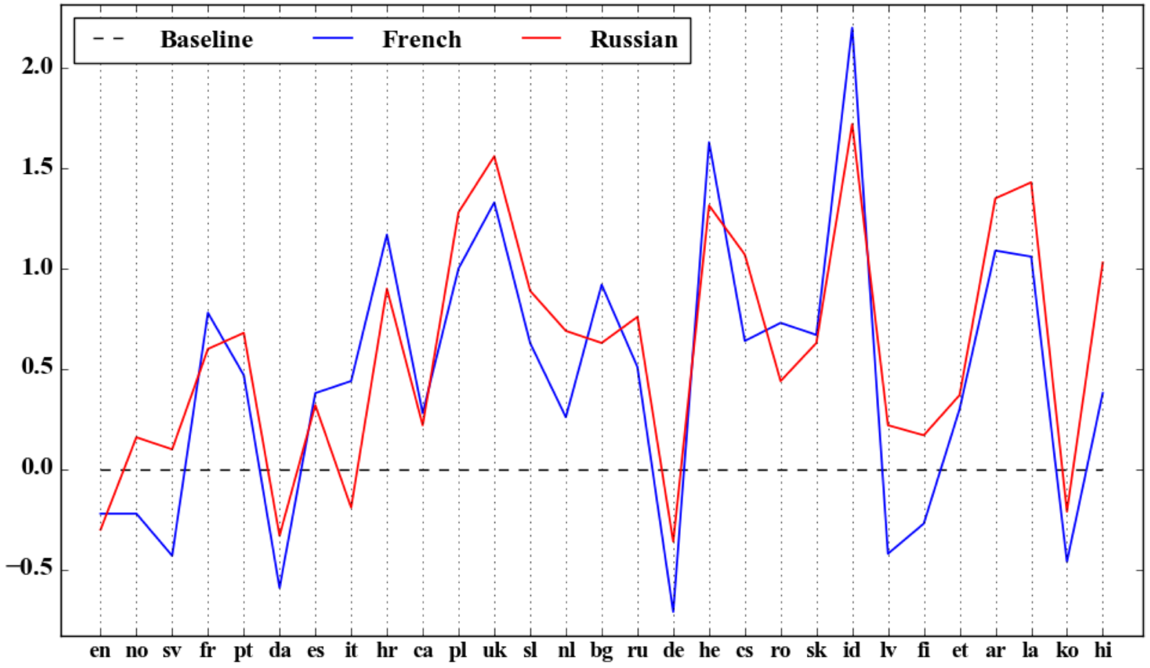
Language Families	Languages
Afro-Asiatic	Arabic (ar), Hebrew (he)
Austronesian	Indonesian (id)
IE.Baltic	Latvian (lv)
IE.Germanic	Danish (da), Dutch (nl), English (en), German (de), Norwegian (no), Swedish (sv)
IE.Indic	Hindi (hi)
IE.Latin	Latin (la)
IE.Romance	Catalan (ca), French (fr), Italian (it), Portuguese (pt), Romanian (ro), Spanish (es)
IE.Slavic	Bulgarian (bg), Croatian (hr), Czech (cs), Polish (pl), Russian (ru), Slovak (sk), Slovenian (sl), Ukrainian (uk)
Korean	Korean (ko)
Uralic	Estonian (et), Finnish (fi)

Table 1: The selected 29 languages for experiments from UD v2.2 (Nivre et al., 2018).

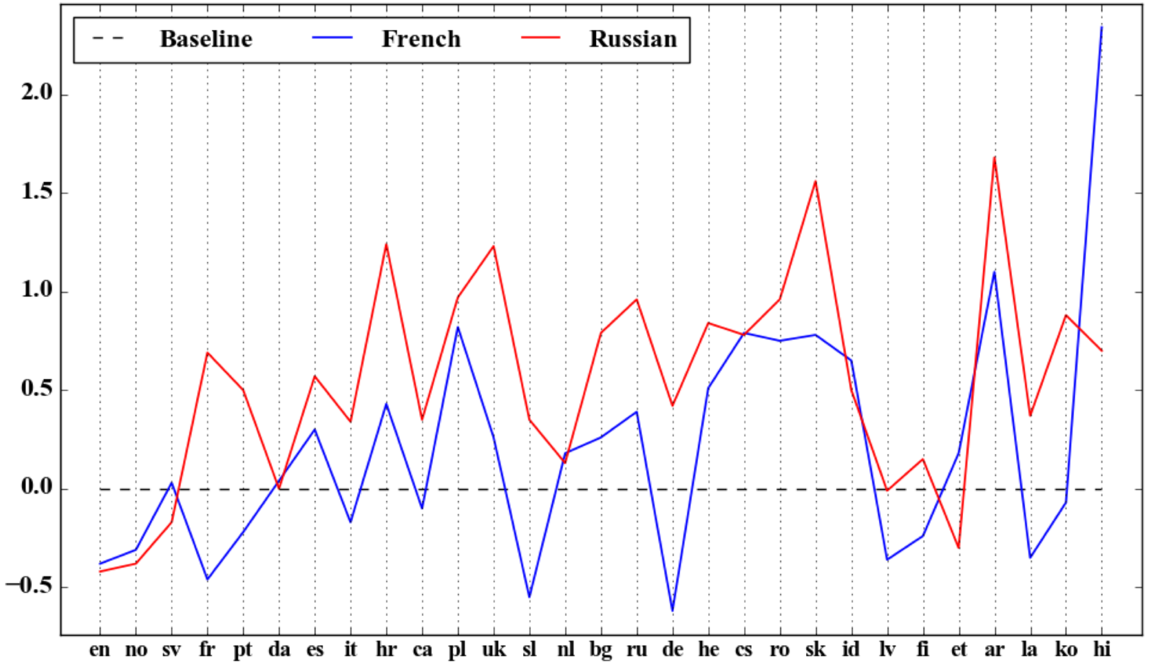
Impact of Adversarial Training

Self-attentive-Graph Parser

Multilingual Word Embeddings



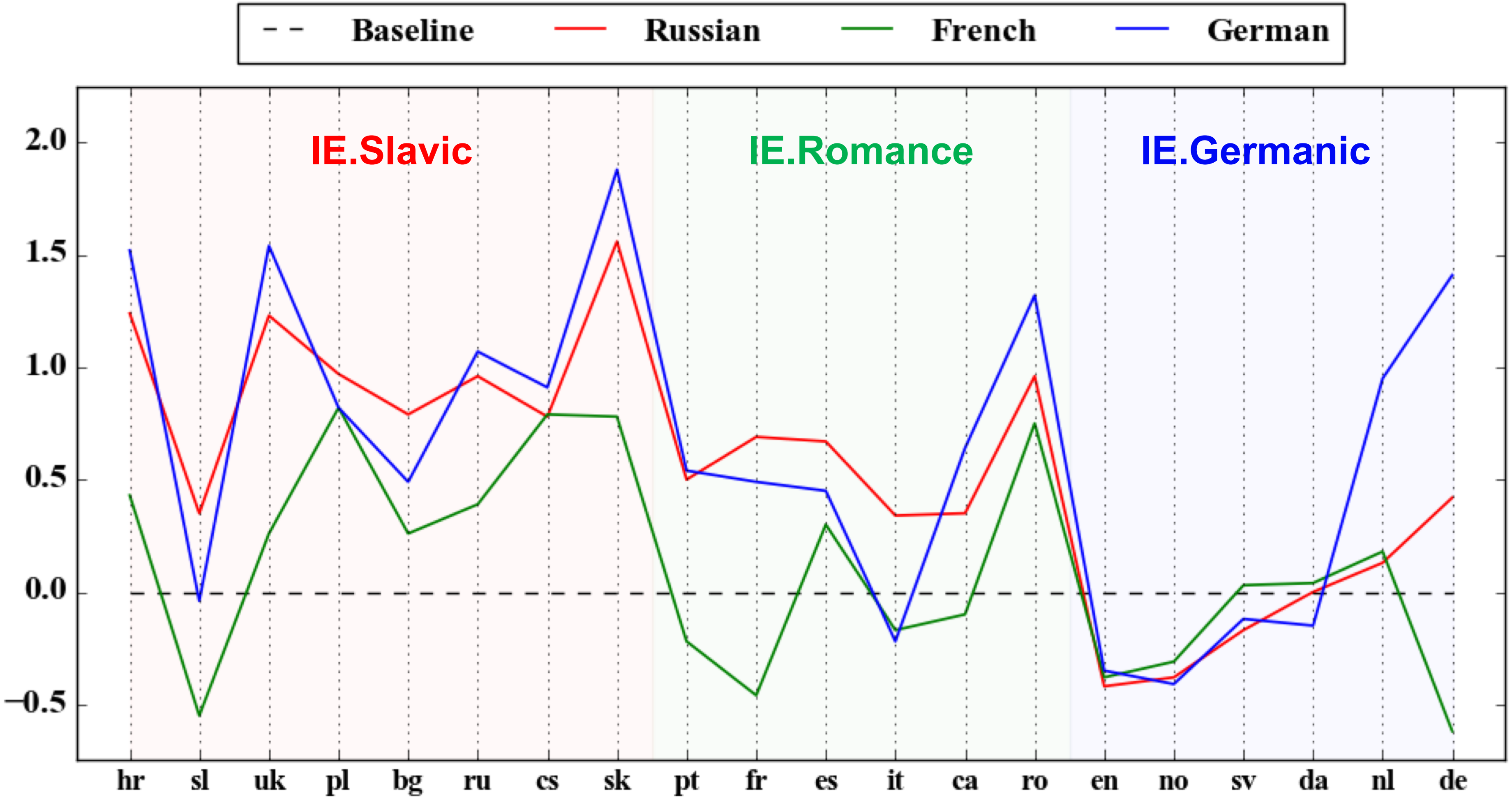
Multilingual BERT



x-axis = language labels
y-axis = performance_diff (model_trained_with_aux_lang, model_trained_on_src_lang)

Impact on Language Families

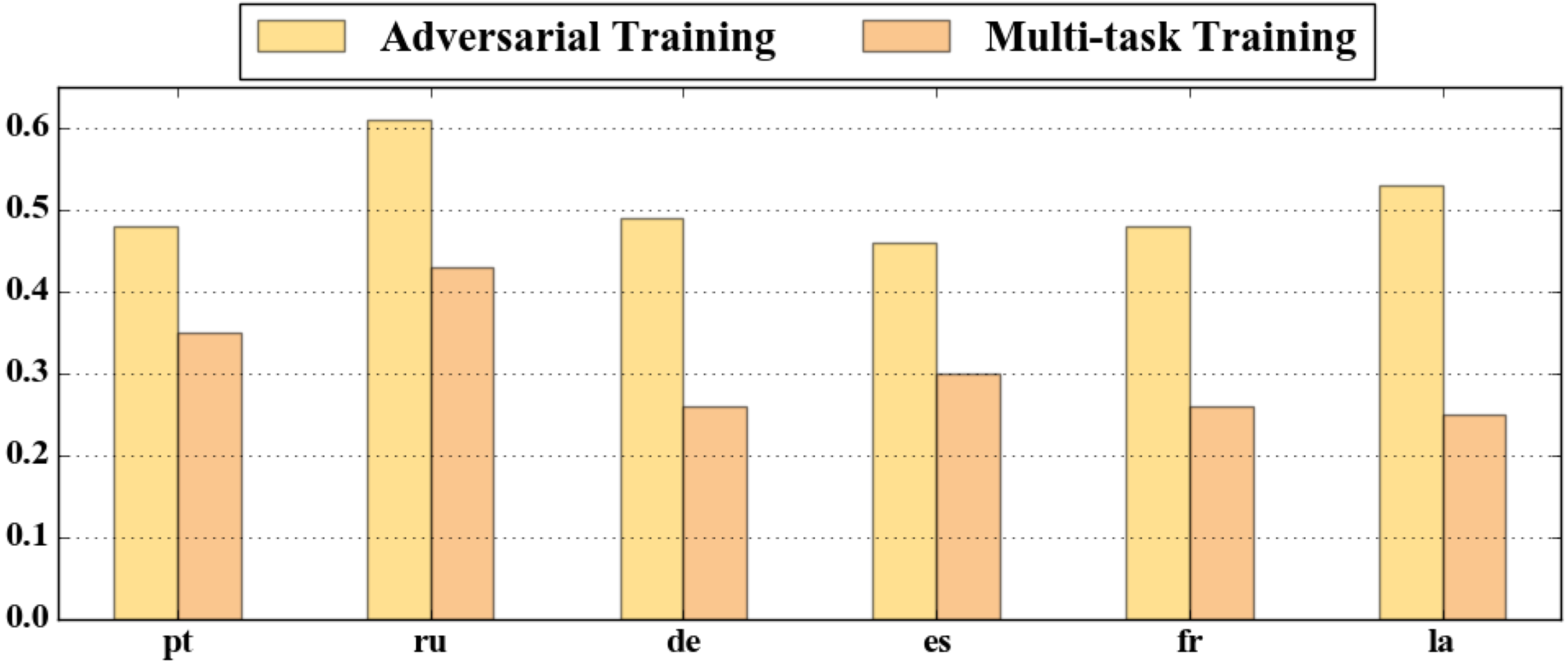
Self-attentive-Graph Parser



Adversarial training (AT) vs. Multi-task learning (MTL)

Adversarial Training (AT)	Multi-task Learning (MTL)
Encoder is trained to maximize the classifier's loss	Encoder is trained to minimize the classifier's loss
Learns to avoid distinguishable language features	Learns to retain distinguishable language features
Observes same amount of data	Observes same amount of data

Adversarial training (AT) vs. Multi-task learning (MTL)

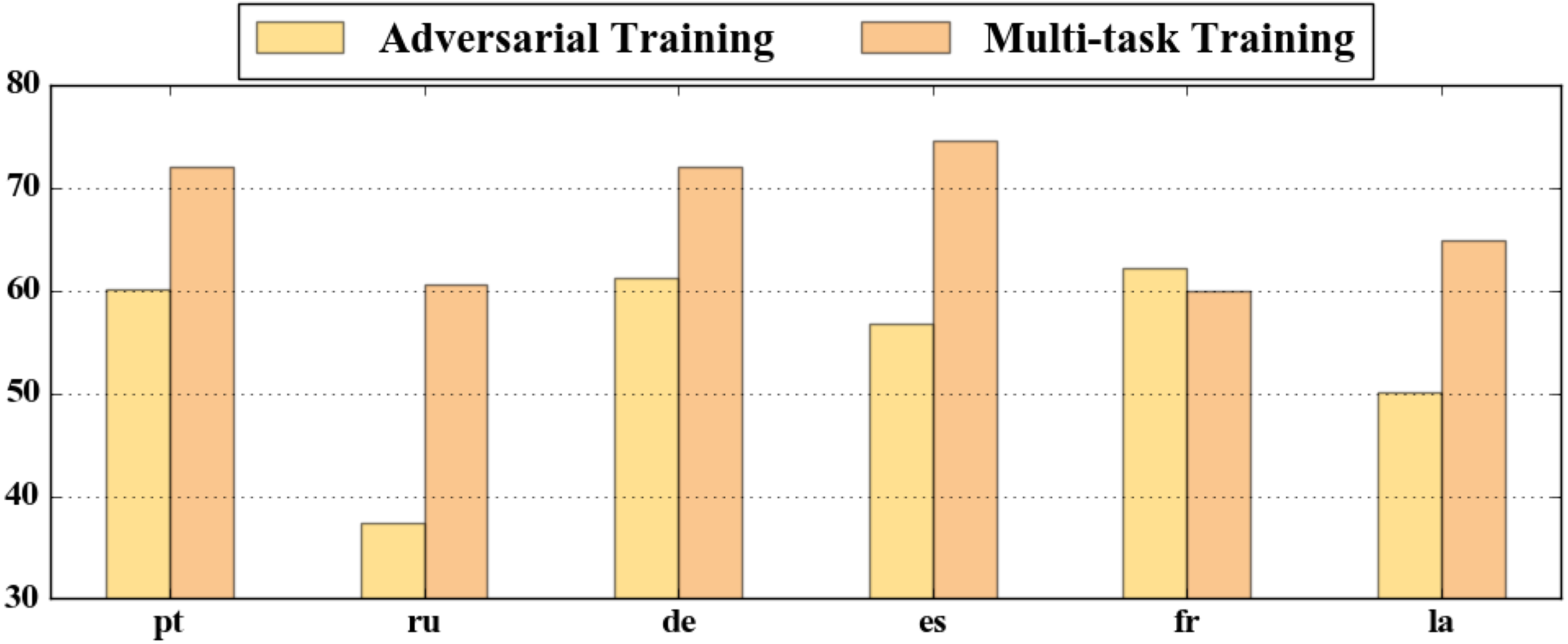


x-axis = auxiliary language labels
y-axis = performance_diff (model_trained_with_aux_lang, model_trained_on_src_lang)

Language Test

- Tests how much encoders retain language information
- A multi-layer perceptron (MLP) as a 7-way classifier
  - Trained on the source and six auxiliary languages
- Metric: accuracy

Adversarial training (AT) vs. Multi-task learning (MTL) – Language Test



x-axis = auxiliary language labels
y-axis = language test performance

Source Code is Publicly Available
https://github.com/wasiahmad/cross_lingual_parsing

Conclusion

- Utilization of Unlabeled Language Corpora
  - Improves cross-lingual dependency parsing
- Adversarial training
  - To learn language-agnostic representation
- Comprehensive empirical study
- Future work
  - Multi-source Transfer Parsing

References

[Ahmad et al., 2019] On Difficulties of Cross-Lingual Transfer with Order Differences: A Case Study on Dependency Parsing. Wasi Ahmad*, Zhisong Zhang*, Xuezhe Ma, Eduard Hovy, Kai-Wei Chang, and Nanyun Peng. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019.

[Smith et al., 2017] Samuel L Smith, David HP Turban, Steven Hamblin, and Nils Y Hammerla. 2017. Offline bilingual word vectors, orthogonal transformations and the inverted softmax. International Conference on Learning Representations.

[Bojanowski et al., 2017] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5:135–146.

[Devlin et al., 2017] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019.

Thank You