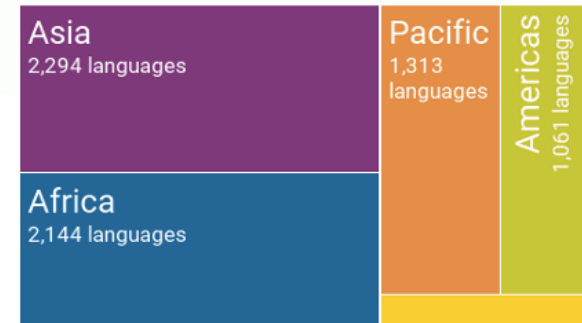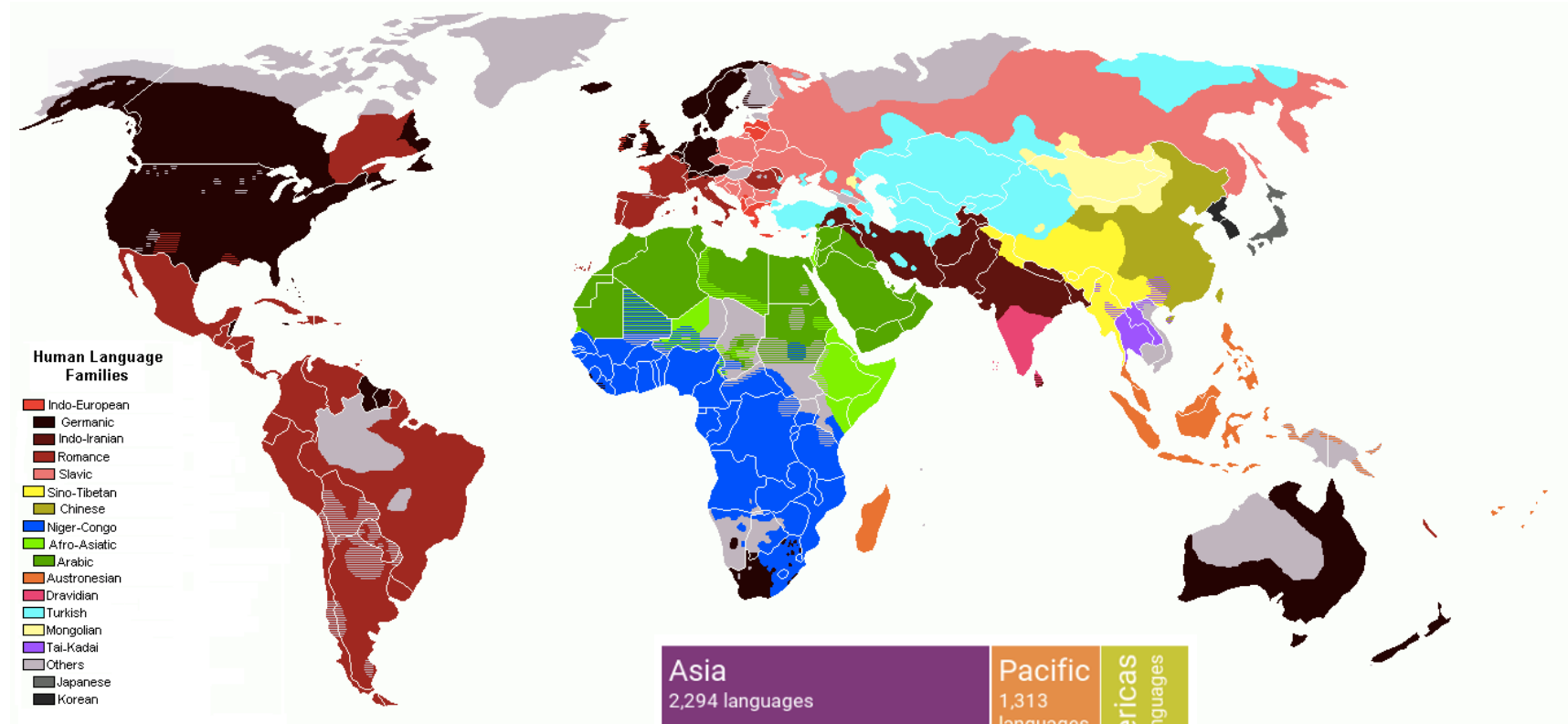# Inducing Language-agnostic Multilingual Representations

Wei Zhao

TU Darmstadt

Joint work with Steffen Eger, Johannes Bjerva and Isabelle Augenstein

NLLG

This work is in submission.

Languages of the World

Human Language Families

- Indo-European
- Germanic
- Indo-Iranian
- Romance
- Slavic
- Sino-Tibetan
- Chinese
- Niger-Congo
- Afro-Asiatic
- Arabic
- Austronesian
- Dravidian
- Turkish
- Mongolian
- Tai-Kadai
- Others
- Japanese
- Korean

Asia 2,294 languages

Pacific 1,313 languages

Americas 1,061 languages

Africa 2,144 languages

Ethnologue

Cross-lingual NLP Systems (1)

[CLS] 图 宾 根 的 机 器 学 习 暑 期 学 校 很 棒 [SEP]

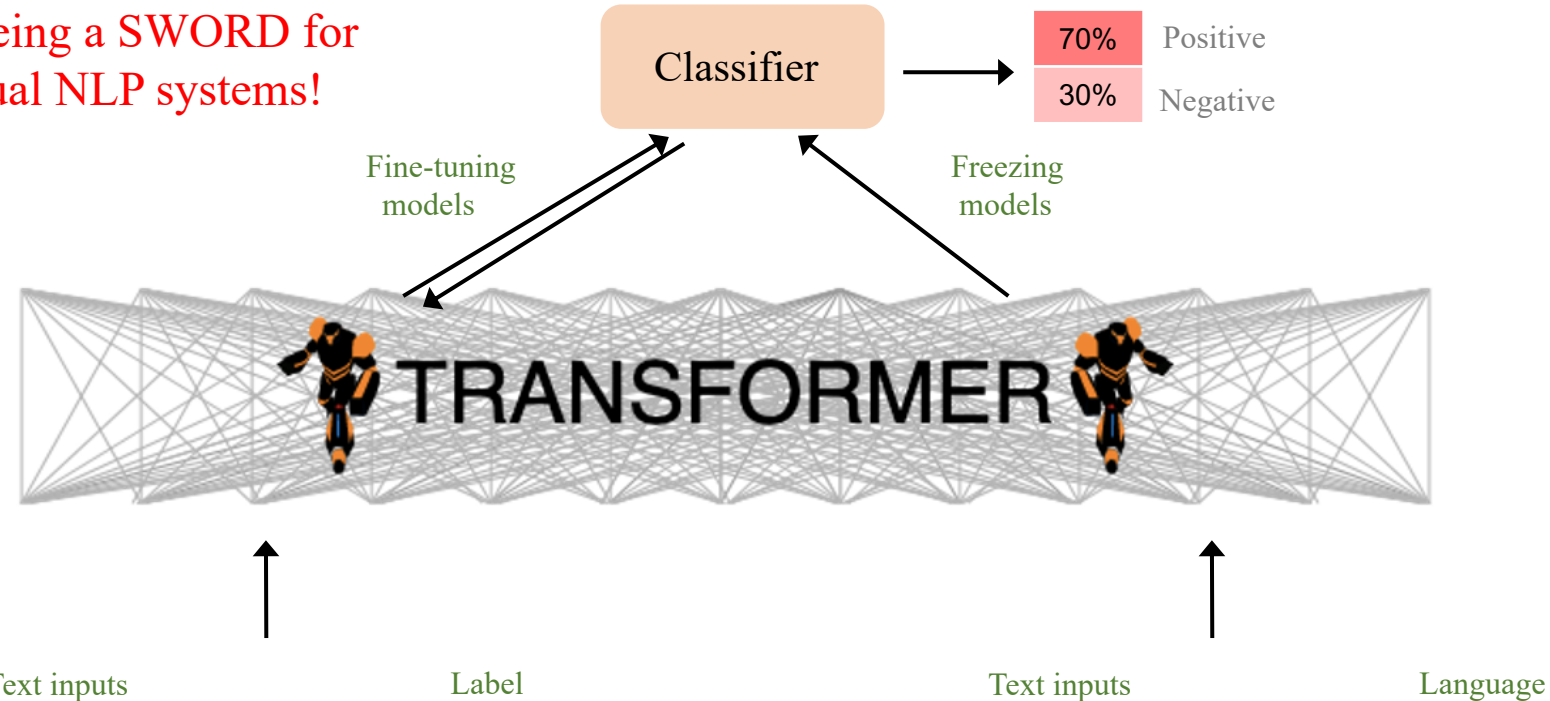[CLS] Die Sommerschule für maschinelles Lernen in Tübingen ist großartig [SEP]

[CLS] The machine learning summer school in Tübingen is great [SEP]

[CLS] The *[MASK]* learning summer school *[MASK]* Tübingen is great [SEP]

TRANSFORMER

machine   in

Multilingual-BERT, XLM-R and others 3

Cross-lingual NLP Systems (2)

Transfer learning being a SWORD for building cross-lingual NLP systems!

Classifier

| 70% | Positive |
| 30% | Negative |

Fine-tuning models

Freezing models

Models

TRANSFORMER

Text inputs | Label

Deep engineering is boring	Negative
Reinforcement learning is slow to train	Negative
Learning to reason is great	Positive

Text inputs | Language

深度工程是无聊的	Chinese
Deep Engineering ist langweilig	German
深いエンジニアリングは退屈です	Japanese

Datasets

Cross-lingual Transfer still in the Stone (1)

XNLI Text Classification

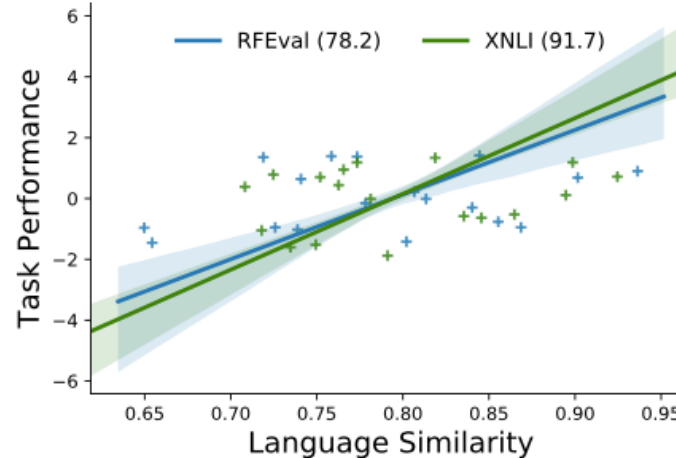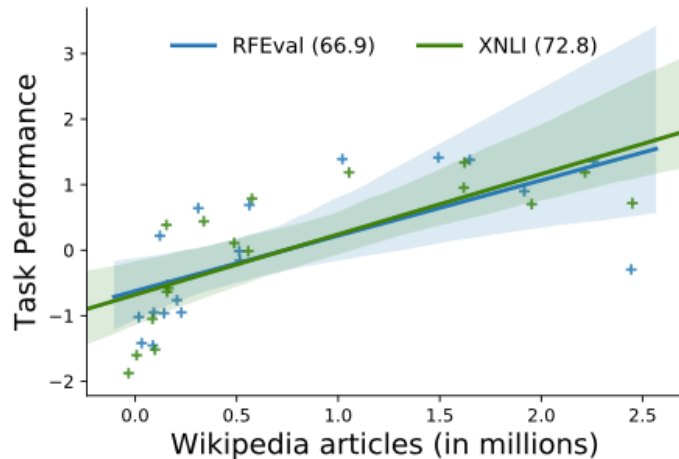| Language	Premise / Hypothesis	Genre	Label
English	You don't have to stay there. You can leave.	Face-To-Face	Entailment
French	La figure 4 montre la courbe d'offre des services de partage de travaux. Les services de partage de travaux ont une offre variable.	Government	Entailment
Spanish	Y se estremeció con el recuerdo. El pensamiento sobre el acontecimiento hizo su estremecimiento.	Fiction	Entailment
German	Während der Depression war es die ärmste Gegend, kurz vor dem Hungertod. Die Weltwirtschaftskrise dauerte mehr als zehn Jahre an.	Travel	Neutral
Swahili	Ni silaha ya plastiki ya moja kwa moja inayopiga risasi. Inadumu zaidi kuliko silaha ya chuma.	Telephone	Neutral
Russian	И мы занимаемся этим уже на протяжении 85 лет. Мы только начали этим заниматься.	Letters	Contradiction
Chinese	让我告诉你，美国人最终如何看待你作为独立顾问的表现。 美国人完全不知道您是独立律师。	Slate	Contradiction

Cross-lingual Transfer still in the Stone (1)

Reference-less evaluation

Case	No.	Source and System Text Pairs	Human	MoverScore
de-en	1.	: Gleichzeitig schließt Kingfisher landesweit 60 B&Q-Filialen : At the same time, Kingfisher will close 60 B & Q stores nationwide	38	125
	2.	: Hewlett-Packard soll bis zu 30.000 Stellen abbauen : Hewlett-Packard will reduce jobs up to 30.000	119	39
	3.	: Serbien ist nach ungarischer Meinung ein "sicheres Drittland". : According to Hungarian view, Serbia is a "safe third country."	23	96
	4.	:And it's from this perspective that I will watch him die. : Experts believe that the Black Friday in November has put the brakes on spending	73	147
	5.	: Und aus dieser Perspektive werde ich ihn sterben sehen. : And from this perspective, I will see him die.	37	111

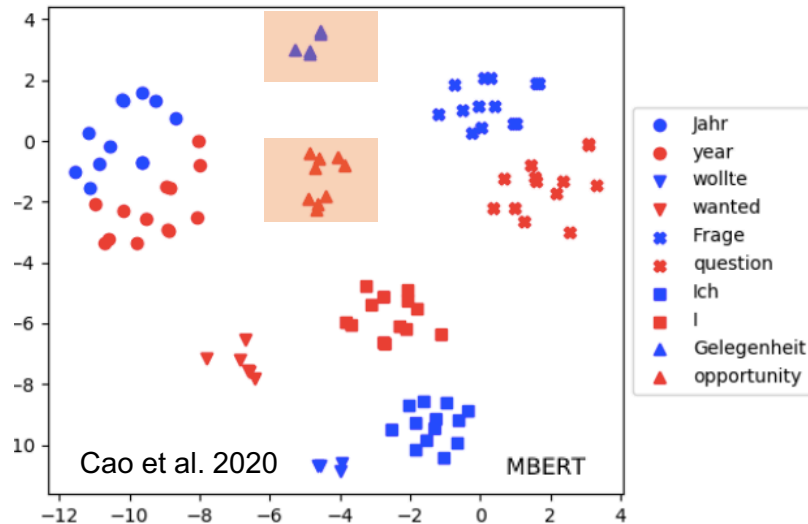# Cross-lingual Transfer still in the Stone (2)
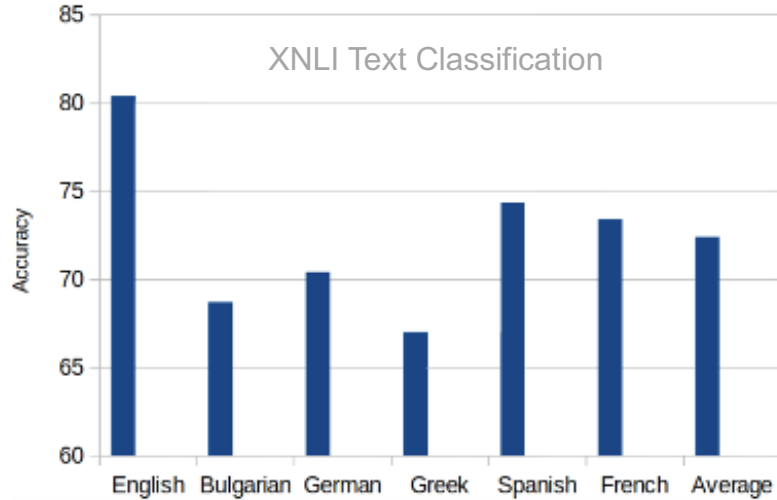
XNLI Text Classification

- Two language constrains are crucial to the success of cross-lingual transfer:
  - The language similarity between source (English) and target languages
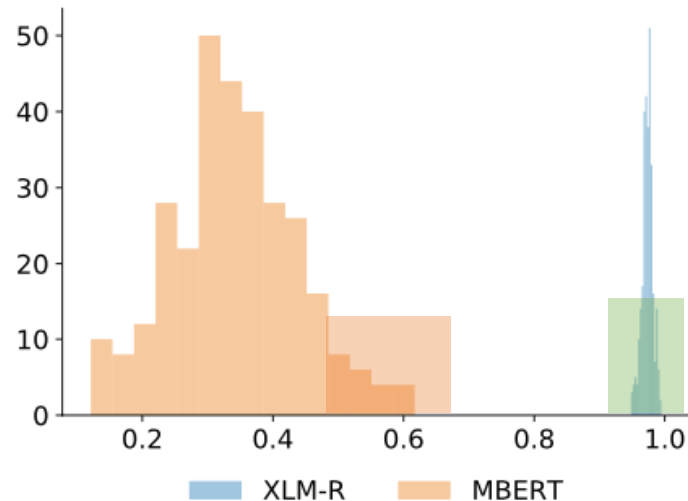 - The size of pre-training corpora for target languages

But many languages are resource-lean and distant to English!

Cross-lingual Transfer still in the Stone (2)

XNLI Text Classification

- Two language constrains are crucial to the success of cross-lingual transfer:
  - The language similarity between source (English) and target languages
  - The size of pre-training corpora for target languages

But many languages are resource-lean and distant to English!

Cross-lingual Transfer still in the Stone (3)

Cao et al. 2020

MBERT

XLM-R   MBERT

Bottom: Histograms of cosine similarity scores of German and English word pairs.

- Transfer to high resource languages that are similar to English is still not perfect, because:

  - The vector spaces of these languages are misaligned

  - Poor discriminative ability, i.e., failure to distinguish matched from random word pairs

Drawing the Sword from the Stone

Libovicky et al. 2019

- Average embeddings of languages (language centroids) within the same language family lie closely in the vector space due to their similar typological features


WALS.info
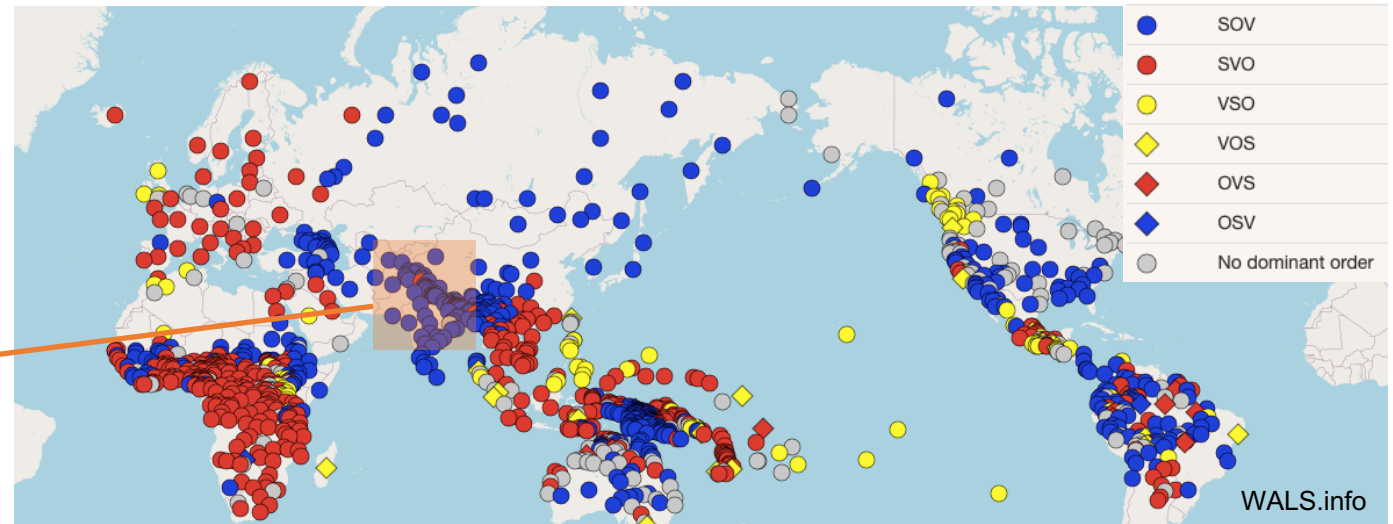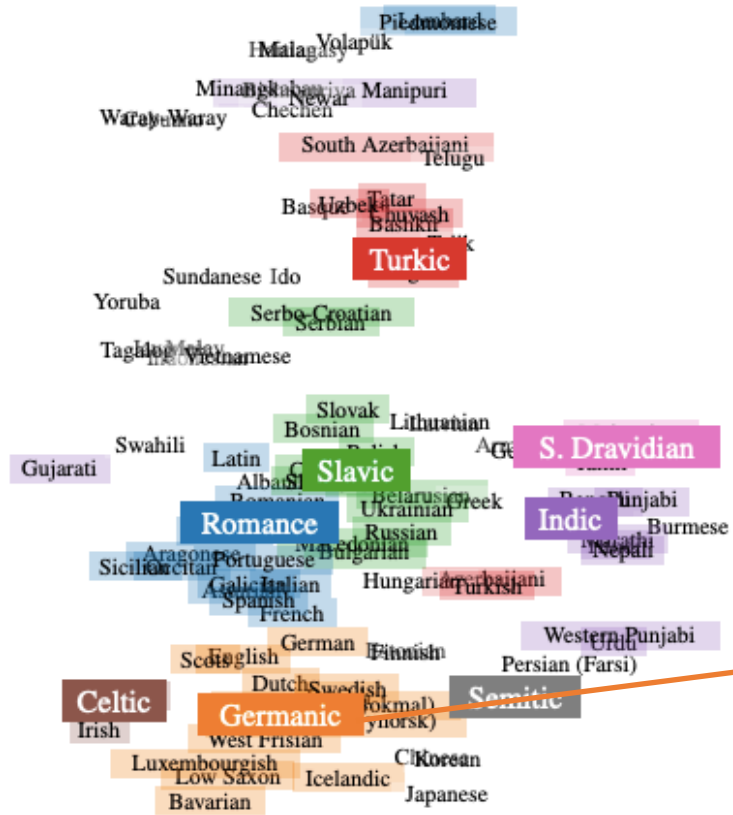
- Language-specific representations make cross-lingual transfer difficult to work!
- Is it possible to produce language-agnostic representations, via removing language identity signals?

Trilogy to Generate Language-agnostic Representations (this work)

- Re-align the vector spaces of target languages to a pivot source language
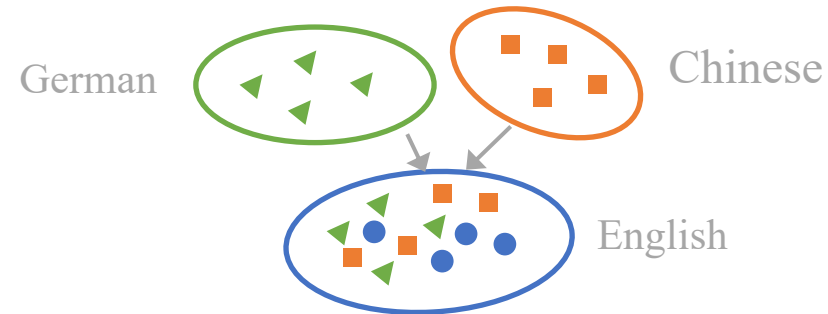
- Remove language-specific means and variances
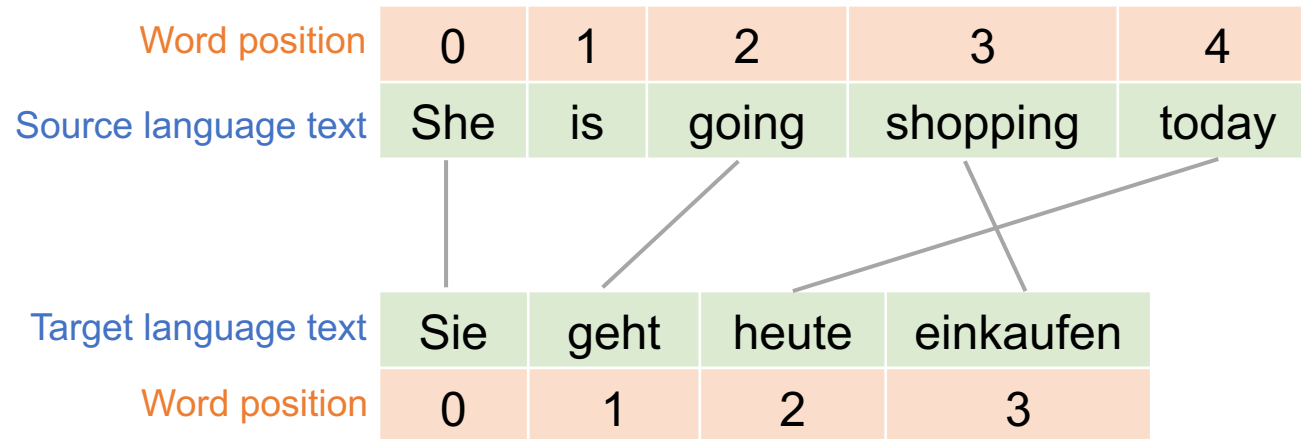- Normalize text inputs by decreasing the linguistic difference of languages
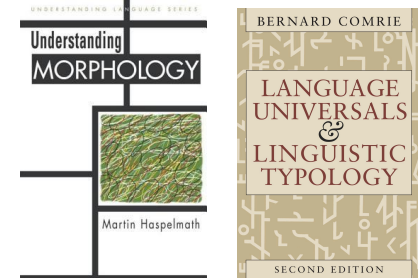
Re-aligning the Vector Spaces of Languages

Parallel Corpus $\quad C = \{[s^1, t^1], \dots, (s^n, t^n)]$

Word Alignment $\quad a = \{[i_1, j_1], \dots, (i_m, j_m)]$

Word Embedding $\quad f_\theta(i, s)$

$$a(s, t) = \{(0, 0), (2, 1), (3, 3), (4, 2)\}$$

Re-aligning the Vector Spaces of Languages

- Find a cross-lingual encoder that has the minimum distance of the embeddings w.r.t. mutual word translations and produces meaningful embeddings (avoid tricky solutions)

$$\min_{\theta} \sum_{(s,t)\in C} \sum_{(i,j)\in a(s,t)} \| f_\theta(i,s) - f_\theta(j,t) \|_2^2 + \sum_{t\in C} \sum_{i} \| f_\theta(i,t) - f_0(i,t) \|_2^2$$

- The vector spaces of target languages could be re-aligned all together to a pivot source language.

13

Normalizing Multilingual Representations

- Find multilingual embeddings that remove language-specific means and variances

$$\hat{f}_\theta(i,s) = \frac{f_\theta(i,s) - u}{\sqrt{\sigma^2 + \epsilon}}$$

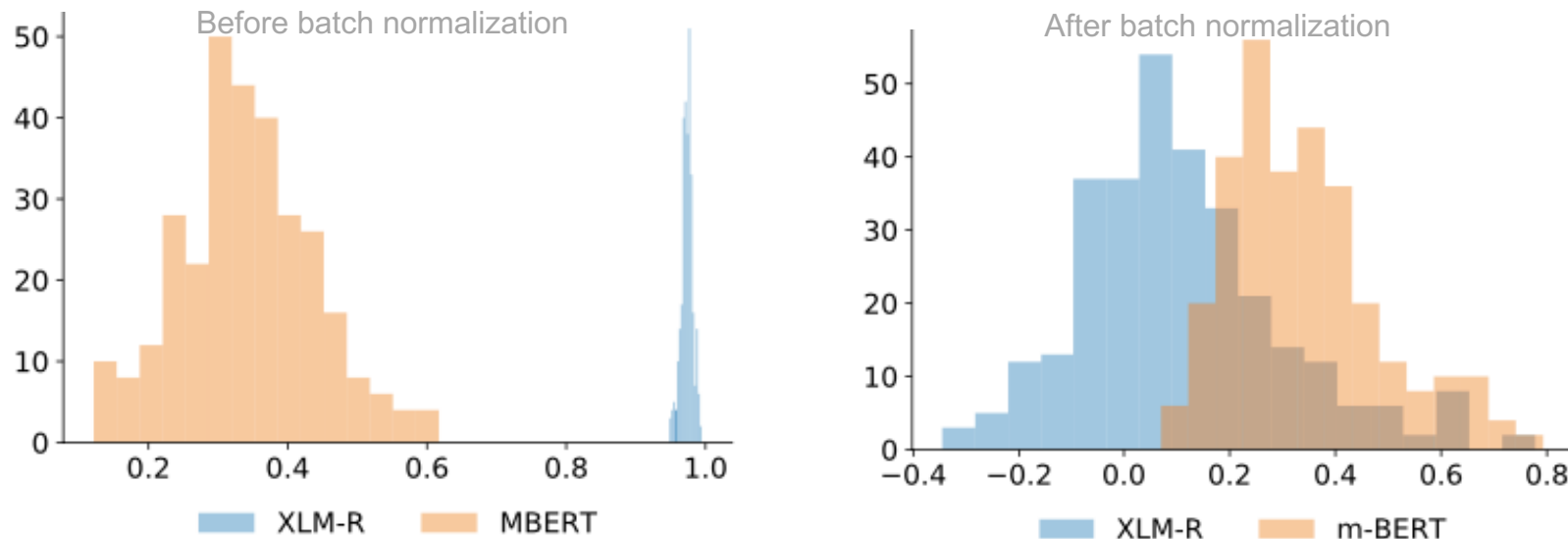- Increase the discriminative ability as a by-product

Figure 1 and 2: Histograms of cosine similarity scores of German and English word pairs.

Normalizing Text Inputs (but it's tricky)

- Different morphological contradictions in languages
  - In some languages, prepositions and articles are often contradicted

| French (target) | de il → del | of it |
| | à le → au | to the |

| German (target) | zu dem → zum | to the |

  - Remove word contradictions in target languages

Normalizing Text Inputs (but it's tricky)

- Different word orders in languages

Source	English	SVO	Adjective-Noun
Target	French	SVO	Noun-Adjective
	Italian	SVO	Noun-Adjective
	Spanish	SVO	Noun-Adjective
	Hindi	SOV	Adjective-Noun
	Turkish	SOV	Adjective-Noun

- Reverse the word order in target languages

How Language-agnostic is Multilingual BERT?
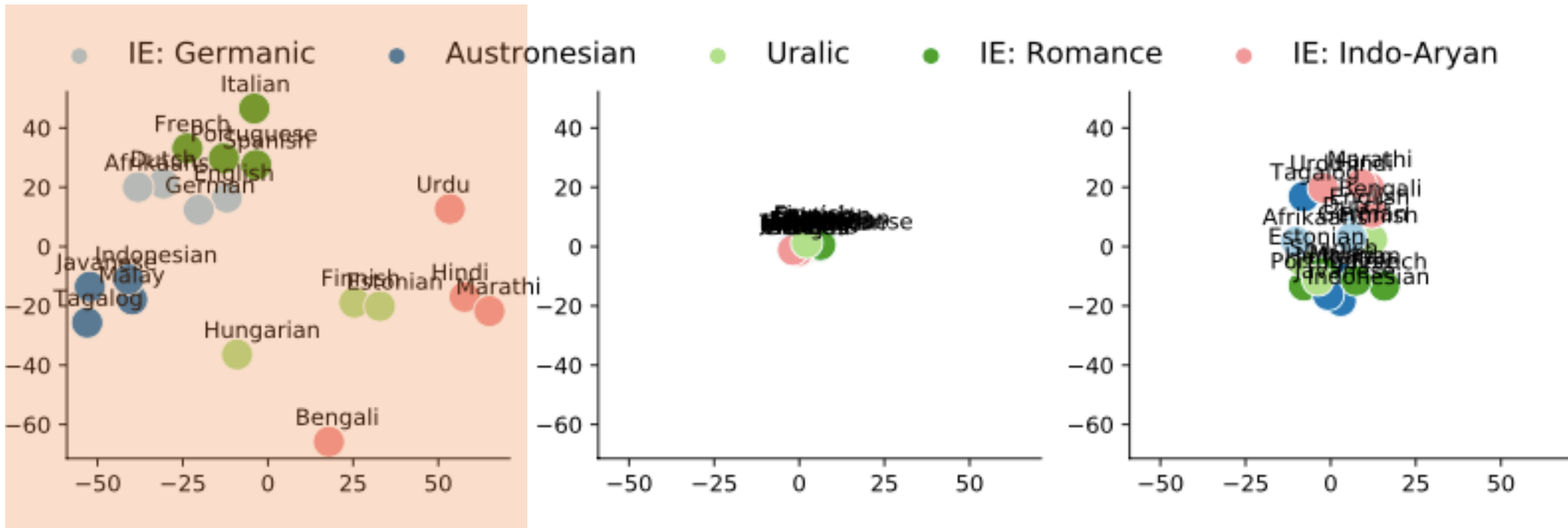
- Language centroids within the same language family lie closely in the vector space (Left).

# How Language-agnostic is Multilingual BERT?
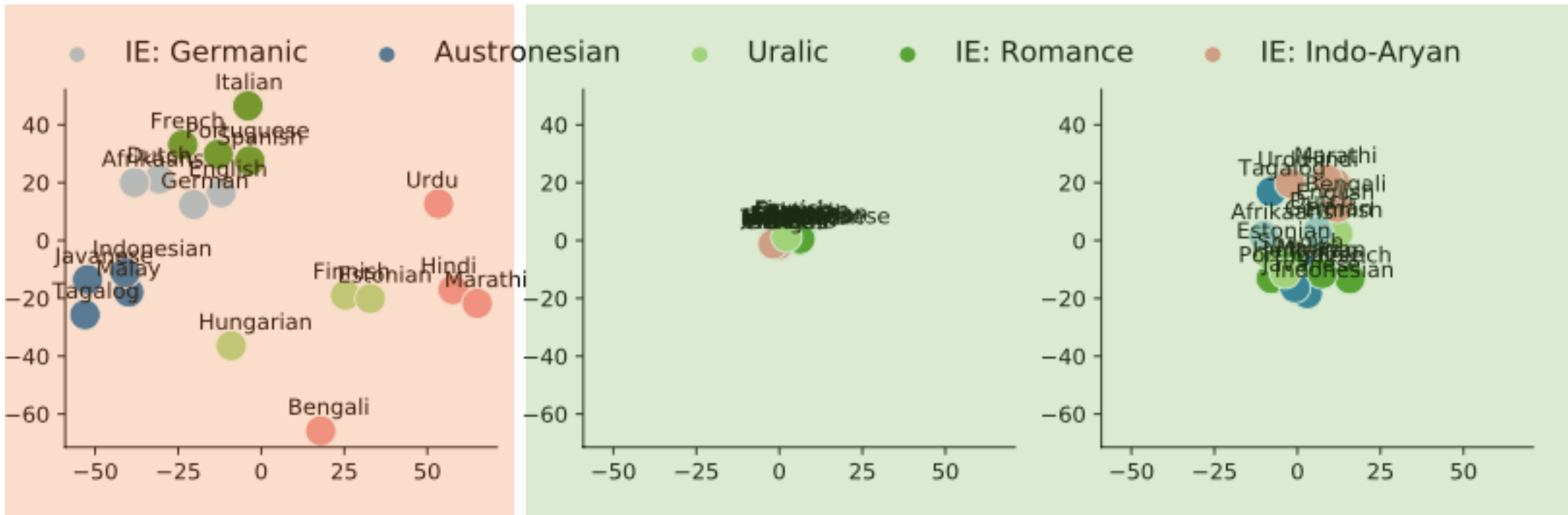
- Language centroids within the same language family lie closely in the vector space (Left).

- Language identity signals are diminished in the re-aligned (Middle) and normalized vector space (Right).
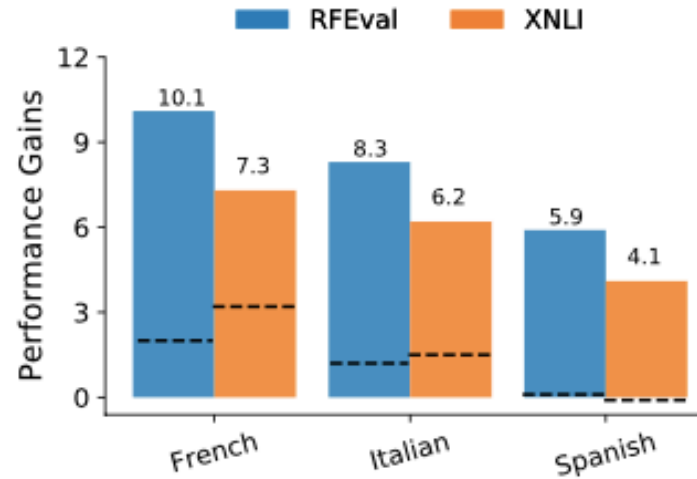
18

Performance Gains across Languages

The similarity between source and target language
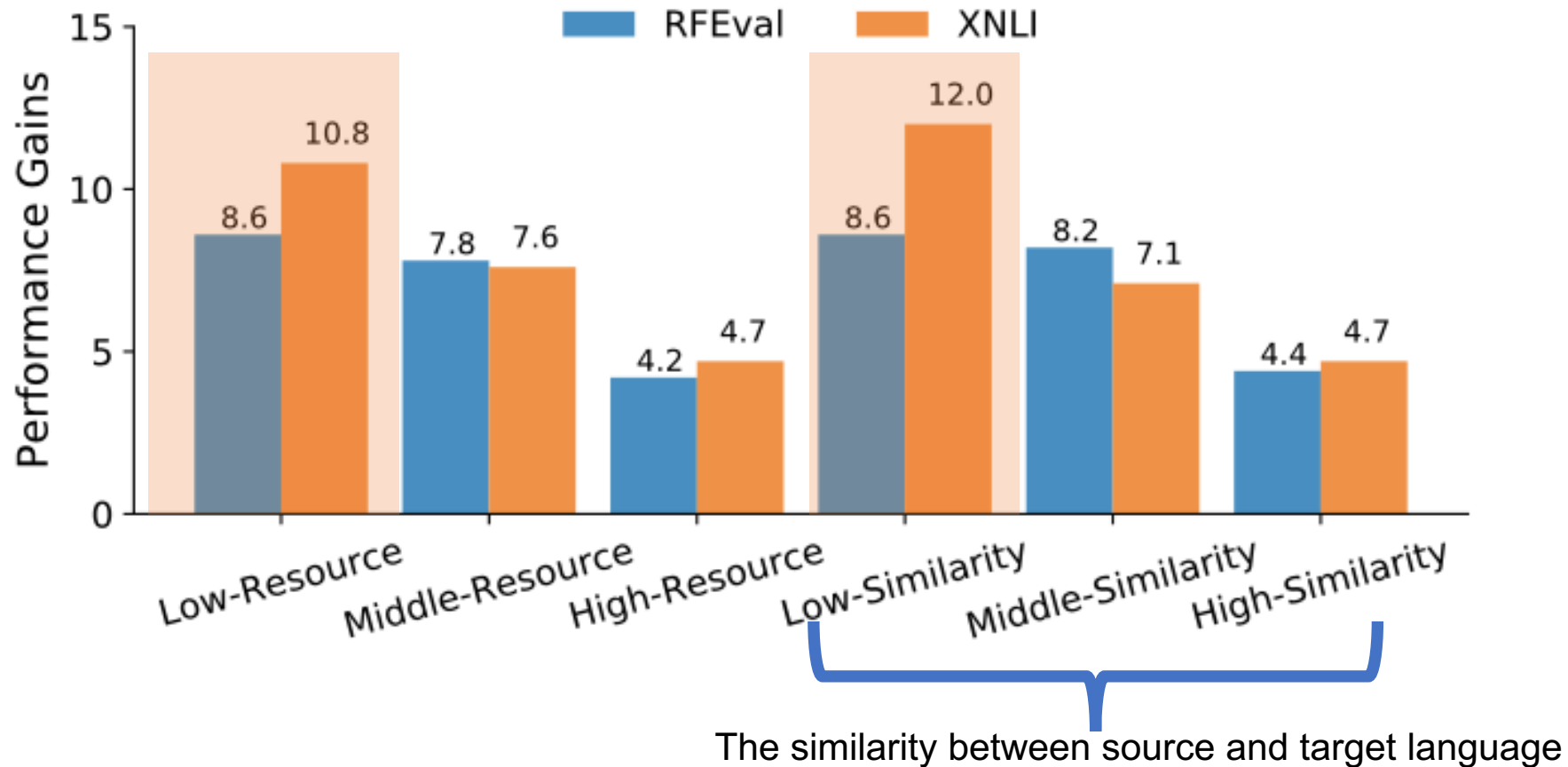
- Cross-lingual transfer achieves the biggest gains on target languages that are resource-lean or distant to English.

- The language constrains can be addressed to some degree.
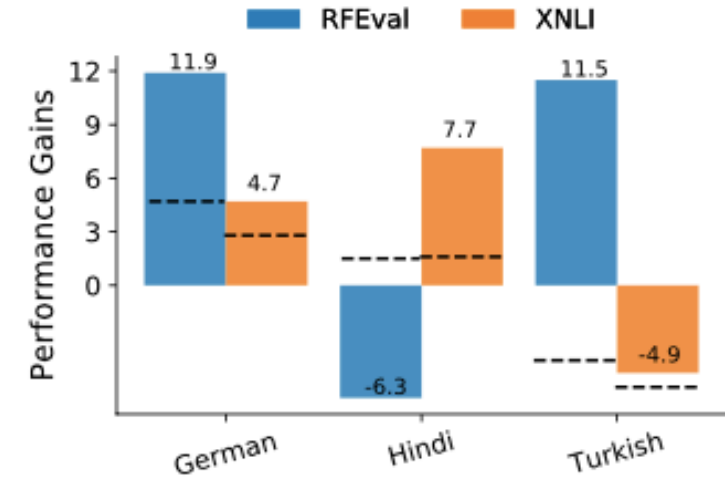
Impact of Text Manipulation
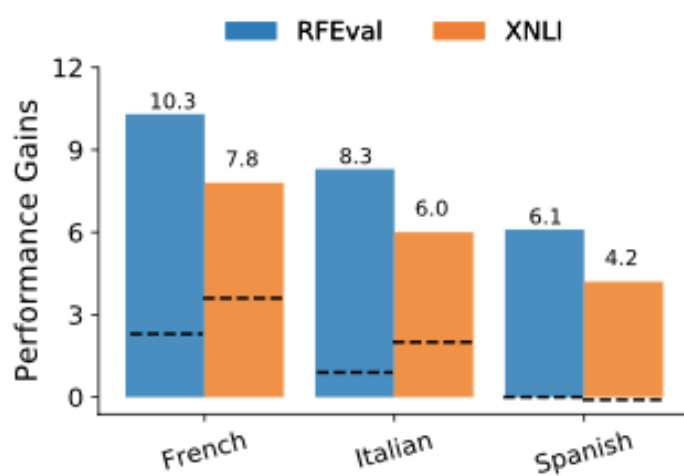
(a) Removing contractions
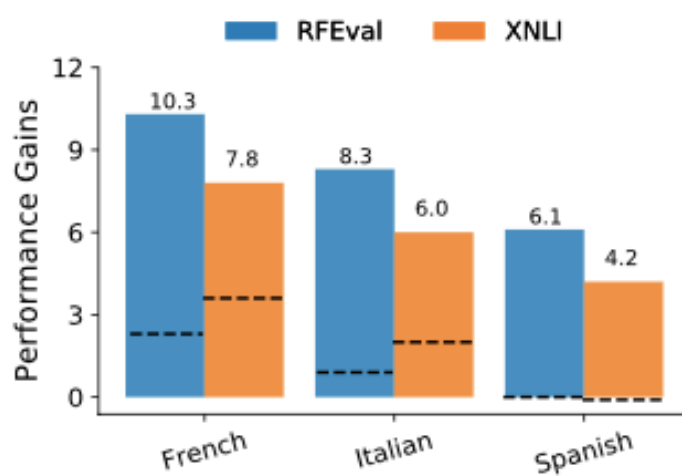
(b) Reversing adjective-noun order
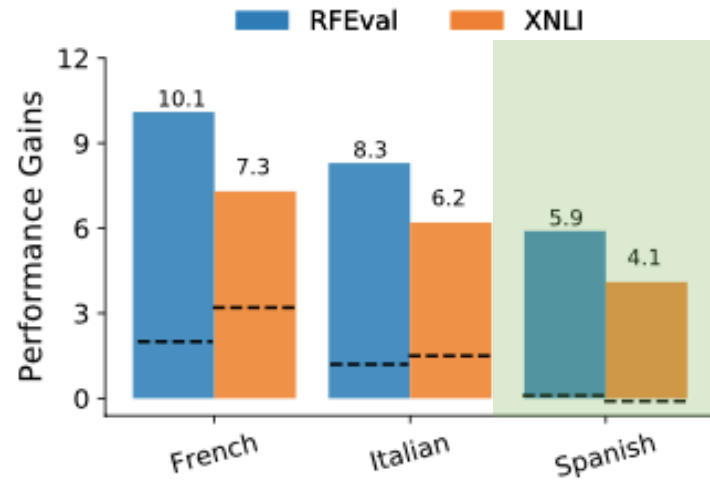
(c) Reversing object-verb order

Full bar denotes the gain from TEXT + NORM + ALIGN, dashed line denotes the gain from TEXT only.

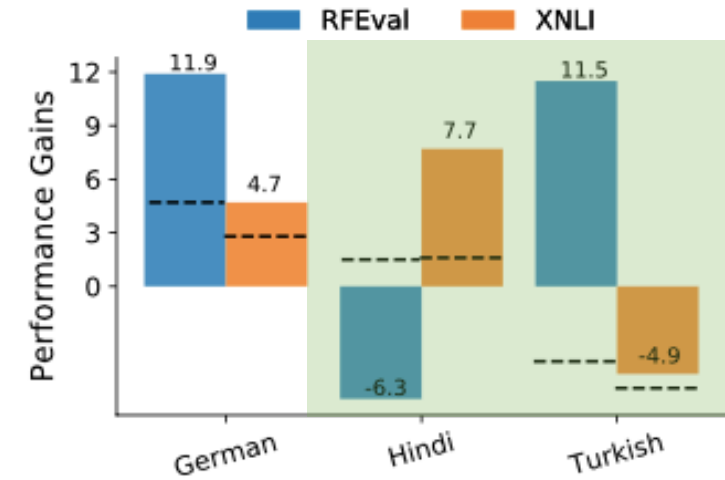- TEXT, NORM and ALIGN are complementary

Impact of Text Manipulation

(a) Removing contractions

(b) Reversing adjective-noun order

(c) Reversing object-verb order

Full bar denotes the gain from TEXT + NORM + ALIGN, dashed line denotes the gain from TEXT only.

- TEXT, NORM and ALIGN are complementary
- The improvements across trial languages are not consistent, because the frequencies of linguistic phenomena in these languages might be different in end-tasks

21

Conclusions

- The success of cross-lingual systems is hindered by two language constrains

- Our language-agnostic representations can remove language identity signals and substantially decrease the cross-lingual transfer gap

- Allow for reference-less machine translation and (cross-lingual) summarization evaluation

- Further discussion: zhao@aiphes.tu-darmstadt.de

fin.