Multilingual BERT has an accent

Evaluating English influences on fluency in multilingual models

Isabel Papadimitriou*, Kezia Lopez*, Dan Jurafsky
Multilingual Models

- Impressive, cross-language results!

Multilingual BERT → Multilingual Task → Metric

Often translated
Multilingual Models

- Impressive, cross-language results!

This paper: **fluency evaluation**, the effect of English syntactic structure on co-trained languages.
Main idea: focus on constructions that are in free variation

Pick a construction in a language which has two forms: one which is parallel to English, one which is not
Main idea: focus on constructions that are in free variation

Pick a construction in a language which has two forms: one which is parallel to English, one which is not

**Spanish:** Pronoun drop

1. Entonces ella *toma* la bandera de la revolución
   So she *take.3SG* the flag of DET revolution

2. Escribió numerosas obras de historia
   write.3SG.PST numerous works of history
Main idea: focus on constructions that are in free variation

Pick a construction in a language which has two forms: one which is parallel to English, one which is not

**Spanish:** Pronoun drop

(1) Entonces ella toma la bandera de la revolución
So she take.3SG the flag of DET revolution

(2) Escribió numerosas obras de historia
write.3SG.PST numerous works of history

**Greek:** Subject-Verb order

(3) Ο πρώτος αγώνας έληξε με σκορ 3:2
The first match finished with score 3:2

(4) Το σκορ του αγώνα άνοιξε ο Γουέν Ρούνι
The score of match opened DET Wayne Rooney
Does mBERT prefer the English-like construction?

Multilingual BERT

Test model

$P(\text{prefers parallel})$

Monolingual model

control model

$P(\text{prefers parallel})$
Method:

Getting an approximation for $P(\text{prefers parallel})$
Data

- Pick out sentences from UD to make the **parallel** and **different** corpora

**Spanish with pronoun**
N = 283

**Spanish prodrop**
N = 2,656

**Greek Subject-Verb**
N = 1,446

**Greek Verb-Subject**
N = 425
Data

- Pick out sentences from UD to make the **parallel** and **different** corpora

Spanish with pronoun  
N = 283

Spanish prodrop  
N = 2,656

Greek Subject-Verb  
N = 1,446

Greek Verb-Subject  
N = 425

- Important that we don’t construct data – we’re looking for fluency
Multilingual BERT

Monolingual model

Spanish with pronoun
N = 283

Spanish prodrop
N = 2,656
Multilingual BERT

Monolingual model

Spanish with pronoun
N = 283

Spanish prodrop
N = 2,656

$P(\text{parallel}|\text{multilingual})$

$P(\text{different}|\text{multilingual})$
Multilingual BERT

Monolingual model

Spanish with pronoun
N = 283

Spanish prodrop
N = 2,656

\( P(\text{parallel} | \text{multilingual}) \)

\( P(\text{different} | \text{multilingual}) \)
Multilingual BERT

Monolingual model

Spanish with pronoun
N = 283

Spanish prodrop
N = 2,656

$P(\text{parallel}|\text{multilingual})$

$P(\text{different}|\text{multilingual})$

$P(\text{parallel}|\text{monolingual})$

$P(\text{different}|\text{monolingual})$
\[ P(\text{prefers parallel}|\text{model}) = \frac{P(\text{parallel}|\text{model})}{P(\text{different}|\text{model})} \]

How to approximate probability of construction?
\[ P(\text{prefers parallel} | \text{model}) = \frac{P(\text{parallel} | \text{model})}{P(\text{different} | \text{model})} \]

How to approximate probability of construction?

- Pick one word in each construction to represent it

**Spanish:** the verb

Entonces **ella** toma la bandera de la revolución

**Escribió** numerosas obras de historia

**Greek:** first word of subject/verb

Ο πρώτος **αγώνας** έληξε με σκορ 3:2

Στις 3_Σεπτεμβρίου **ξέσπασε** επανάσταση
Results

\[ P(\text{prefers parallel}|\text{model}) = \frac{P(\text{parallel}|\text{model})}{P(\text{different}|\text{model})} \]

Spanish pro-drop

Greek subject-verb order

![Graph showing comparisons between monolingual and multilingual models for Spanish pro-drop and Greek subject-verb order]
Fluency evaluation
Fluency evaluation

- The effect probably won’t influence downstream, non-generation performance very much
Fluency evaluation

- The effect probably won’t influence downstream, non-generation performance very much
- And there are so many other things that we’re not looking at
Fluency evaluation

- The effect probably won’t influence downstream, non-generation performance very much
- And there are so many other things that we’re not looking at
  - Subtleties in lexical choice, grammatical choice, discourse…
Fluency evaluation

- The effect probably won’t influence downstream, non-generation performance very much
- And there are so many other things that we’re not looking at
  - Subtleties in lexical choice, grammatical choice, discourse…
- Entanglement of fluency with domain variation, translationese
Fluency evaluation

- The effect probably won’t influence downstream, non-generation performance very much
- And there are so many other things that we’re not looking at
  - Subtleties in lexical choice, grammatical choice, discourse…
- Entanglement of fluency with domain variation, translationese
- Future work: evaluations that don’t use a monolingual model
Fluency evaluation

- The effect probably won’t influence downstream, non-generation performance very much
- And there are so many other things that we’re not looking at
  - Subtleties in lexical choice, grammatical choice, discourse…
- Entanglement of fluency with domain variation, translationese
- Future work: evaluations that don’t use a monolingual model
  - Preserving stimuli naturalness in a controlled experiment
Fluency evaluation

- The effect probably won’t influence downstream, non-generation performance very much.
- And there are so many other things that we’re not looking at:
  - Subtleties in lexical choice, grammatical choice, discourse…
- Entanglement of fluency with domain variation, translationese.
- Future work: evaluations that don’t use a monolingual model:
  - Preserving stimuli naturalness in a controlled experiment.

Thanks!