Chère maison or *maison chère*?
Transformer-based Prediction of Adjective Placement in French

Position of attributive adjectives in French

ma <u>chère</u> maison = 🏠 ❤️ vs. *ma maison <u>chère</u>* = 🏠 💰

- (Mostly) anteposed adjectives: *troisième, grand, ...*
- (Mostly) postposed adjectives: *rouge, fabuleux, ...*
- Mobile adjectives (w.r.t. semantics): *ancien, cher, ...*

Our research question

Are transformer-based embeddings sensitive to word order, when positional information is semantically important?

French Transformer-based models

- CamemBERT-{base, large}
- FlauBERT-{small, base, large}-{cased/uncased}

Datasets

frWaC & Universal Dependencies 3.0

Data	Train	Val.	frWaC test	UD test (100%)	UD test (25%)
frWaC	76,164	7,672	7,740	19,437	5,151
frWaC +UD (75%)	91,615	7,672	7,740	-	5,151
UD (75%)	13,905	1,546	7,740	-	5,151

Exp. 1: Finetuning & Classification of attributive adjective position

- Input: Two sentences with different word order in N-ADJ pair (ADJ before = 0, ADJ after = 1)
  *On construit les éléments de **haut niveau**.* - 0
  *On construit les éléments de **niveau haut**.* - 1
- Finetune with different train datasets & domains, two-sentences, one-sentence, also attention masks on context or pair
- Baselines: Log. regression, CNN, frequency
- **Results:**
  - 0.87-0.99 on frWaC
  - 0.97-0.99 on frWaC+UD
  - 0.62-0.99 on UD
  - CamemBERT > FlauBERT, but baselines close
- Masking context only good for CamemBERT, masking N-ADJ good for most!
- Error analysis: few mistakes, some in mobile adjectives, some from parsing

Exp. 2: Testing adjective pretrained embeddings

- **Classification only with adjective embeddings** (and log. regression): moderate, successful only for flaubert-base-uncased
- **MLM probabilities**: higher probability of masked ADJ in original position than opposite (note correlation of anteposition & frequency)

Exp. 3: Human judgments vs models' probabilities

- Dataset of challenging/control sentences:
  1. Adjective/Noun dependents
  2. Fixed expressions
  3. Structural persistence
  4. Blocked/mobile adjectives
- 4 questionnaires, total of 71 human participants
- Correlation of human judgments and model classification probabilities:

Model	1	2	3	4
camembert-base	0.21	-0.19	-0.08	0.47
camembert-large	**0.67**	**0.61**	0.53	0.51
flaubert_small_cased	0.51	-0.03	0.16	**0.78**
flaubert_base_cased	0.52	0.09	0.38	0.71
flaubert_base_uncased	0.40	0.22	**0.63**	0.56
flaubert_large_cased	0.46	0.18	**0.63**	0.47

Discussion

- Easy task... until it's not! Frequency is key
- Finetuning: data-hungry, multiple domains a plus
- Context is crucial and is exploited by models
- But not enough information in the ADJ embedding
- Models vs. Humans:
  - Models have too high probabilities
  - Models prefer postposition, even when wrong
  - Models fall victim to priming?

Eleni Metheniti🟢🟠, Tim Van de Cruys🔵, Wissam Kerkri🟢, Juliette Thuilier🟢, Nabil Hathout🟢
firstname.lastname@{univ-tlse2.fr🟢,irit.fr🟠,kuleuven.be🔵}
CLLE-CNRS🟢, IRIT🟠, KU Leuven🔵