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

Eleni Metheniti, Tim Van de Cruys, Wissam Kerkri, Juliette Thuilier, Nabil Hathout

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

- (Mostly) anteposed adjectives:
  - Ordinals: *troisième* 'third'
  - High-frequency adjectives: *grand* 'big'

- (Mostly) postposed adjectives:
  - Colors: *rouge* 'red'
  - Polysyllabic adjectives: *fabuleux* 'fabulous'

- Mobile adjectives (w.r.t. semantics):
  - e.g. *ancien* 'old, former' or 'ancient'
  - e.g. *cher* 'dear' or 'expensive'

Abeillé and Godard (1999); Thuilier (2013); Benzitoun (2014)

- Are transformer-based embeddings sensitive to word order, when positional information is semantically important?
- Are they already or do they need to learn additionally?
- How do their decisions compare to native speakers?

- Input: Two sentences with different word order in noun-adjective pair (labels: 0 = anteposition, 1 = postposition)

  *On construit les éléments de **haut niveau**.* - 0 - (original)
  *On construit les éléments de **niveau haut**.* - 1 - (permutated)

- Also finetune with one-sentence input (original)
- Also finetune with attention masks:
    - Mask entire context except for noun and adjective
    - Mask noun and adjective, context visible

- Baselines: Logistic regression, CNN, frequency

French Transformer-based models

- CamemBERT (Martin et al., 2020): RoBERTa-based model
  - camembert-base
  - camembert-large

- FlauBERT (Le et al., 2020): BERT-based, XLM elements
  - flaubert-small-cased
  - flaubert-base-cased
  - flaubert-base-uncased
  - flaubert-large-cased

- frWaC (Baroni et al., 2009)
- Universal Dependencies 3.0 (Zeman et al., 2021)
- Different training and test sets for finetuning:

| Source	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

- 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, catastrophic for FlauBERT
- Masking noun-adjective had overall lower results, but not failure
- Error analysis: few mistakes, some in mobile adjectives, some from parsing

· **Classification only with adjective embeddings** (and logistic regression): moderate, successful only circumstantially
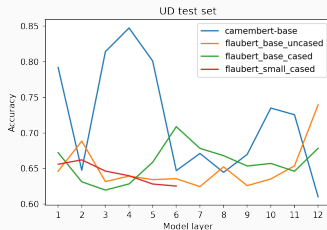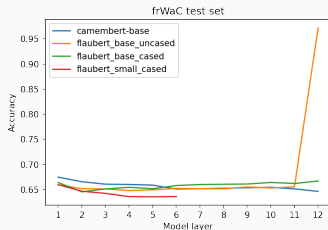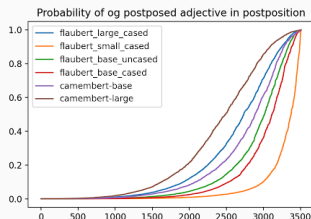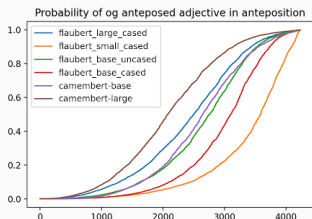
- MLM adjective probabilities:

  Probability of *haut* in the masked position:
  On construit les éléments de [MASK] niveau. (original)
  On construit les éléments de niveau [MASK]. (permutated)

- Higher probability of masked adjective in original position than opposite, postposed adjectives preferred postposition more (note correlation of anteposition & frequency)

Exp. 3: Human judgments vs models' probabilities

- Same two-sentence setup as Experiment 1

- Dataset of challenging/control sentences:
    1. Adjective/Noun dependents
    2. Fixed expressions
    3. Structural persistence
    4. Blocked/mobile adjectives

- 3 questionnaires, a 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

- 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 adjective embedding
- Models vs. Humans:
    - Unacceptable mistakes
    - Models have too high probabilities
    - Models prefer postposition, even when wrong

Thank you for your (unmasked) attention!

Code for this paper:
`https://github.com/lenakmeth/word_order`

Selected References

Abeillé, A. and Godard, D. (1999). La position de l'adjectif épithète en français: le poids des mots. *Recherches linguistiques de Vincennes*, (28):9–32.

Baroni, M., Bernardini, S., Ferraresi, A., and Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, 43(3):209–226.

Benzitoun, C. (2014). La place de l'adjectif épithète en français: ce que nous apprennent les corpus oraux. In *SHS Web of Conferences*, volume 8, pages 2333–2348. EDP Sciences.

Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., Allauzen, A., Crabbé, B., Besacier, L., and Schwab, D. (2020). FlauBERT: Unsupervised language model pre-training for French. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 2479–2490, Marseille, France. European Language Resources Association.

Martin, L., Muller, B., Ortiz Suárez, P. J., Dupont, Y., Romary, L., de la Clergerie, É., Seddah, D., and Sagot, B. (2020). CamemBERT: a tasty French language model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7203–7219, Online. Association for Computational Linguistics.

Thuilier, J. (2013). Syntaxe du français parlé vs. écrit: le cas de la position de l'adjectif épithète par rapport au nom. *TIPA. Travaux interdisciplinaires sur la parole et le langage*, (29).

Zeman, D., Nivre, J., Abrams, M., et al. (2021). Universal dependencies 2.9. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.