# Cumulative frequency can explain cognate facilitation in language models

Irene E. Winther[1,2], Yevgen Matusevych[2] and Martin J. Pickering[3]

[1] UKRI Centre for Doctoral Training in Natural Language Processing
[2] School of Informatics, University of Edinburgh
[3] School of Philosophy, Psychology & Language Sciences, University of Edinburgh

## COGNATE FACILITATION EFFECT

**Cognates** Words that share form and meaning across languages. They have a processing advantage relative to non-cognates.

- Dutch-English bilinguals read example (a) faster than (b) (Bultena et al., 2014)
  a. The residents disliked the **winter** (cognate: Dutch *winter*)
  b. The residents disliked the **prison** (control: Dutch *gevangenis*)

## CUMULATIVE FREQUENCY HYPOTHESIS

- Bilinguals encounter cognates more frequently than non-cognates due to their identical form. Cognate effect just a frequency effect in disguise?[2,3]
  - Cognates do not have a special status in the memory
  - Evidence is non-conclusive: some argue for special status instead[4]

**Can a computational model that doesn't assign special status to cognates show the cognate facilitation effect?**

- Cognate processing in a computational language model (LM) to test cumulative frequency hypothesis
  - Can count exact frequencies in input data
  - Cognates and controls treated the same way by model
  - Compare LM predictions to human data

## METHOD

- LSTM-LM[5] trained on 2 languages: first language (L1) Dutch - second language (L2) English
  - Wikipedia-corpora (2M shuffled sentences) – 80/10/10 training/test/valid
- Training conditions to explore
  1. **Language mixing**
     - **NON-MIX** L1 data followed by L2 data
     - **MIX** L1+L2 data shuffled per sentence
  2. **L1:L2 ratio**
     - 75:**25** 75% L1 data - 25% L2 data
     - 50:**50** equal split
  3. **L1 pretraining**
     - **PT** LM trained on L1 (30 epochs) → L1+L2 data (10 epochs)
     - **NPT** LM trained on L1+L2 data (30 epochs)

## EVALUATION

**Cognate effect** Surprisal on cognates vs. controls for sentence stimuli from Bultena et al. (2014) **(plots A & B)**

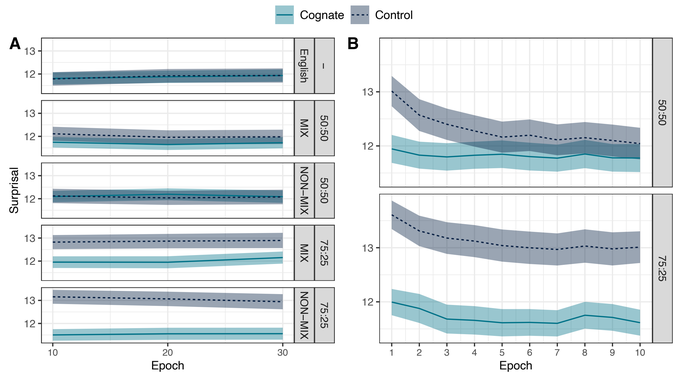
- Surprisal is a measure of processing effort[6] – correlates with human reading times[7]
- Expectation: lower surprisal for cognates

**Linguistic accuracy** Perplexity on L1 and L2 test sets (separately) – compare to monolingual LMs **(plots C & D)**

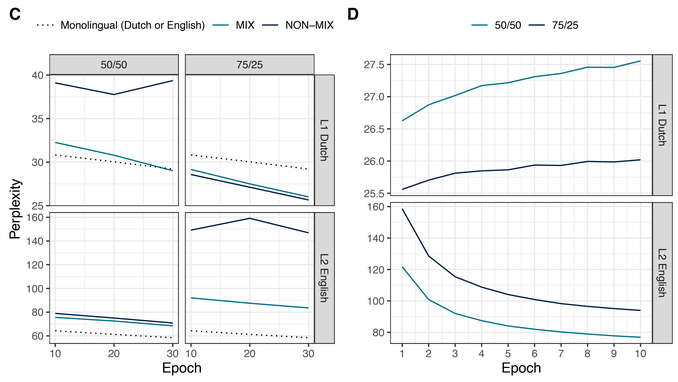
## RESULTS

**NPT models** **PT Models**

**Surprisal – cognates vs. controls**




**Perplexity on L1 and L2 test sets**




## SUMMARY

**Do models exhibit a cognate effect?**

- 2 out of 6 models displayed the effect
  - They have significantly lower surprisal for cognates than controls
  - Common properties of the two models associated with high L2 perplexity
    - Higher exposure to L1 (75:25 language ratio)
    - Presentation of L1 before L2 (NON-MIX or PT)

**Does the magnitude of the effect depend on L2 perplexity?**

- Difference between cognates and controls larger for models with low L2 linguistic accuracy (L2 "proficiency"), measured by perplexity
  - Same trends in humans: larger effect in bilinguals with low L2 proficiency[1,2]

**Does word frequency explain the effect better than the cognate status?**

- Yes, cognate status not significant when frequency included as a predictor
  - Higher frequency of cognates (compared to non-cognates) facilitates their processing in sentences

**Do the results hold for other language pairs?**

- Yes, we ran an identical study using Norwegian-English training data and test stimuli

## CONCLUSION

- **Findings support the cumulative frequency hypothesis**
- **Cognate effect** Lack of exposure to non-cognate words?
  - Cognate words are like high-frequency words for less proficient speakers; non-cognates are like lower-frequency words
  - Differences in exposure to the two types of words decrease with increased proficiency ➔ smaller cognate effects

## REFERENCES

1. Bultena, S., Dijkstra, T., & van Hell, J. G. (2014). Cognate effects in sentence context depend on word class, L2 proficiency, and task. *Quarterly Journal of Experimental Psychology*, *67*, 1214–1241.
2. Pivneva, I., Mercier, J., & Titone, D. (2014). Executive control modulates cross-language lexical activation during L2 reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 787–796.
3. Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: Electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*, *20*, 912–928.
4. van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, *9*, 780–789.
5. van Schijndel, M., & Linzen, T. (2018). A neural model of adaptation in reading. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.
6. Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*.
7. Goodkind, A., & Bicknell, K. (2018). Predictive power of word surprisal for reading times is a linear function of language model quality. In *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)*.

CogSci21