Informativity enhances memory precision in the agreement attraction effect

Weijie Xu, Richard Futrell

University of California, Irvine

28th Architectures and Mechanisms for Language Processing

September 7 - 9, 2022

Introduction

Predictability vs. Informativity

- Predictability affects online processing difficulty
  - > Lower predictability → Higher processing difficulty
  - > Motivates: Surprisal theory [1-2]
- But……
  - > Information theory: Surprisal corresponds to information content [3]
  - > Language is often used to deliver newsworthy message [4]
- We ask: ***What is the role of newsworthy messages in sentence processing?***

Hypothesis

Newsworthy message yields better memory representation

- Maybe, it's about **memory representation**
  - Lower predictability → More cognitive resources in need [5-6]
  - Thus……memory representation is **more robust against noise**
- Test case: **agreement attraction effect**

Test case: Agreement Attraction

The key to the cabinets are…

- Subject-verb agreement on number feature

  e.g. The key was rusty from many years of disuse.

- If there's a distractor N in between

  *e.g. The key to the cabinets {was/*were}…*

  > **Production**

  …more likely to produce the ungrammatical "were" [7]

  > **Comprehension**

  …less likely to notice the ungrammatical "were" [8-9]

current study

Predictions

Memory precision in the agreement attraction effect

*The key to the cabinets {is/*are}…*

Assumption: The encoded memory representation of the target NP has been distorted, resulting in AA effect

Lower predictability on target N → Weaker AA effect **[Experiment 1]**

> More robust representation of target N, less likely to be distorted

Lower predictability on attractor N → Stronger AA effect **[Experiment 2]**

Assumption: Limited pool of cognitive resources

> More resources on attractor N, less resources to maintain target N

> Less robust representation of target N, more likely to be distorted

Experiment 1

Lower predictability on target N → Weaker AA effect?

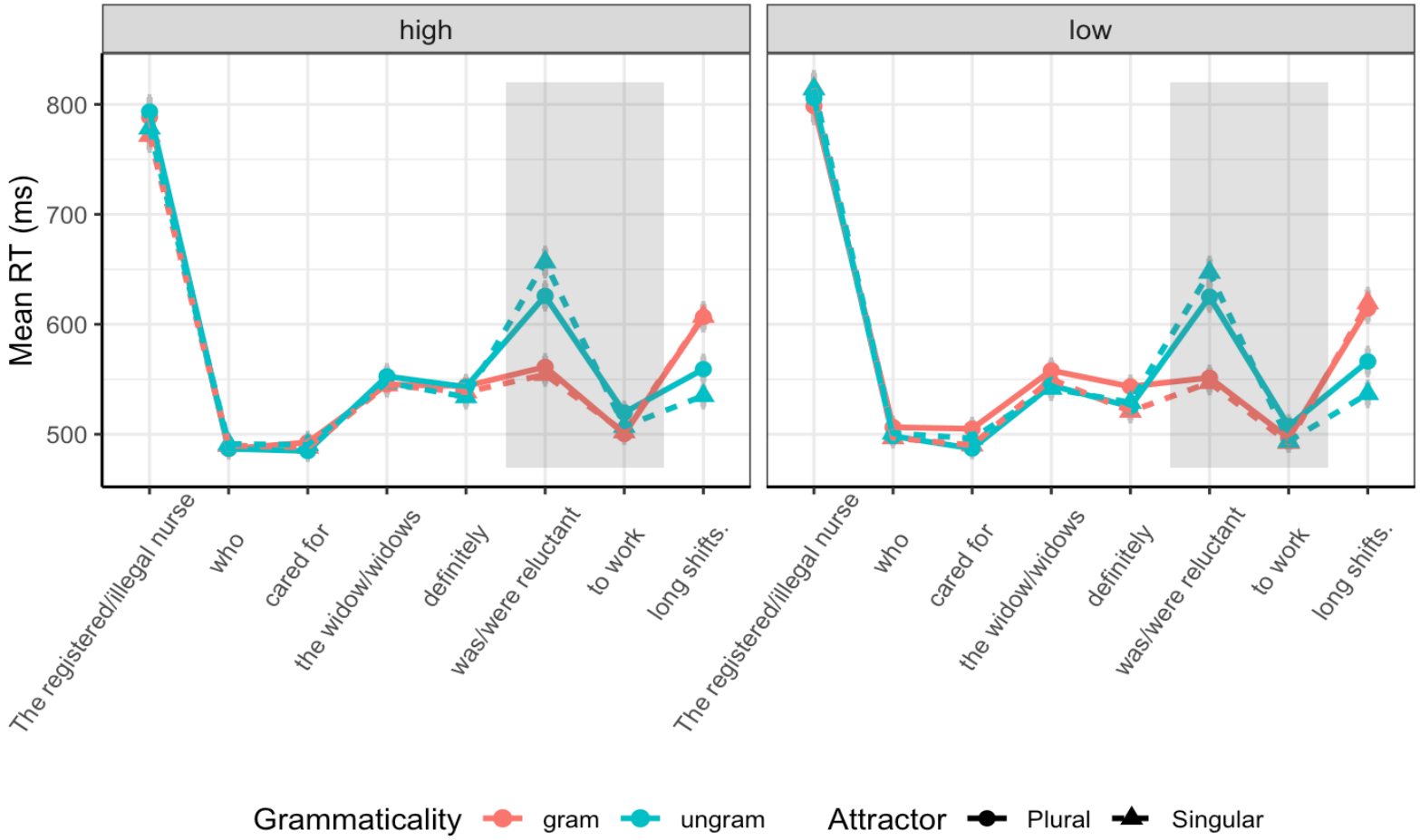
- Self-paced reading (n=194, items=32)
- 2 x 2 x 2 within-subject design
  - **Predictability** of target N
  - **Grammaticality** of the subject-verb agreement
  - **Attractor** NP's number feature

  } AA effect
  - Critical statistics: Predictability x Grammaticality x Attractor
- Sample stimuli
  - [typical target NP] **The registered nurse**/ who/ cared for/ the {widow/widows}/ definitely/ {was/were} reluctant/ to work/ long shifts.
  - [atypical target NP] **The illegal nurse**/ who/ cared for/ the {widow/widows}/ definitely/ {was/were} reluctant/ to work/ long shifts.

Experiment 1

Lower predictability on target N → Weaker AA effect?

- Results 1a: Binary categorization of predictability

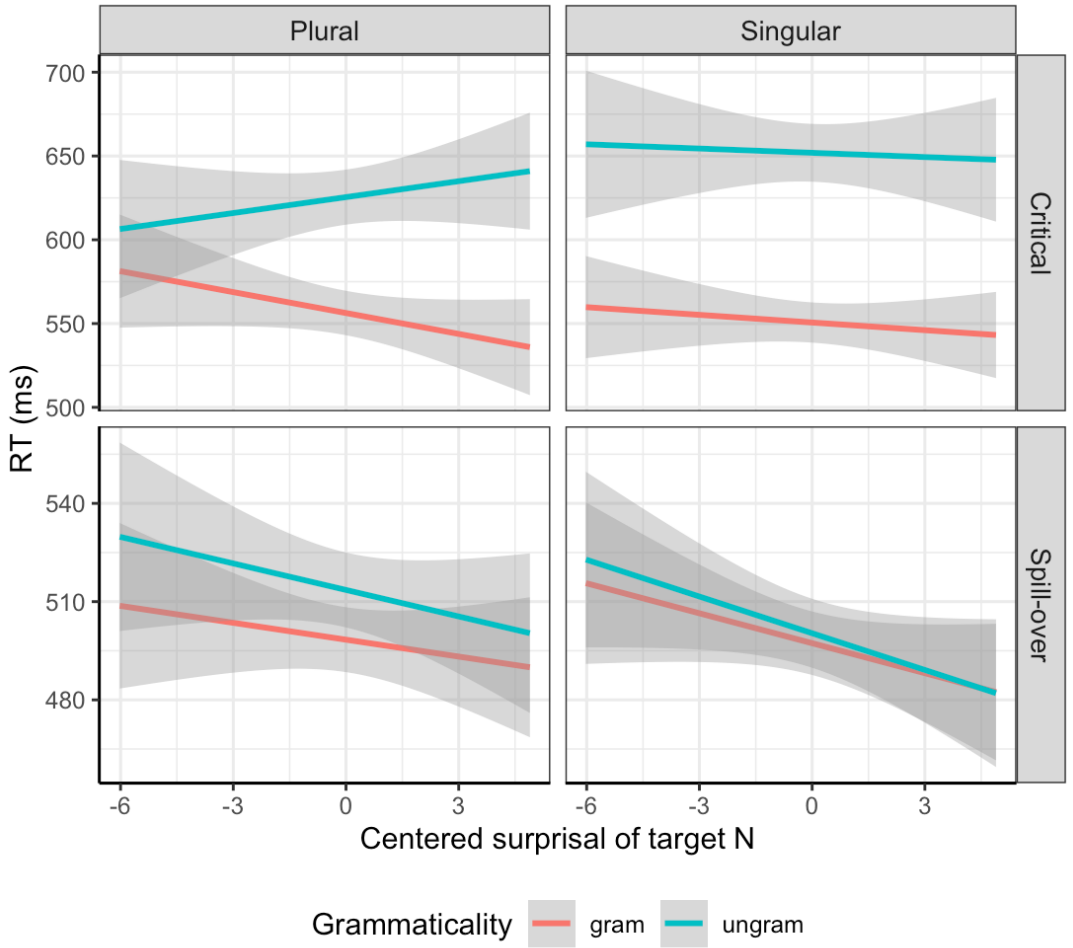


No evidence for the standard Gram x Attractor agreement attraction effect regardless of the predictability condition

Experiment 1

Lower predictability on target N → Weaker AA effect?

- Results 1b: Surprisal of head N in subject NP generated by GPT-3 [10]



- Significant Gram x Surprisal interaction in the critical region for plural attractor
- Such an interaction was not detected for singular attractor
- No effect in the spill-over region

Discussion

Memory precision in the agreement attraction effect

☑ **Lower predictability on target N → Weaker AA effect** **[Experiment 1]**

- > We did observe weaker AA effect if higher surprisal on target N
- > Support: Newsworthy information yields better memory representation

? **Lower predictability on attractor N → Stronger AA effect** **[Experiment 2]**

Experiment 2

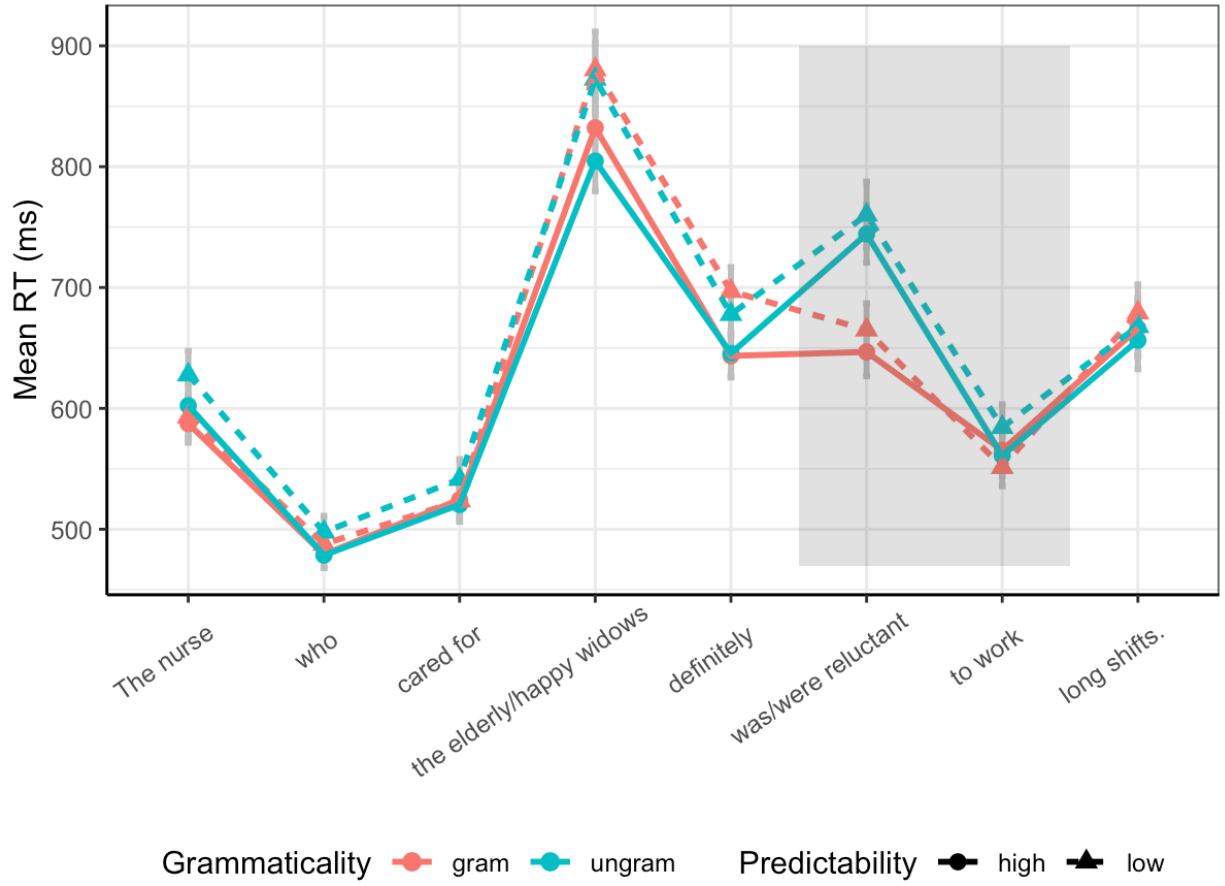
Lower predictability on attractor N → Stronger AA effect?

- A pilot experiment: Self-paced reading (n=60, items=16)
- 2 x 2 within-subject design
  - **Predictability** of attractor N
  - **Grammaticality** of the subject-verb agreement → AA effect
  - Critical statistics: Predictability x Grammaticality
- Sample stimuli
  - [typical target NP] The nurse/ who/ cared for/ **the elderly widows**/ definitely/ {was/were} reluctant/ to work/ long shifts.
  - [atypical target NP] The nurse/ who/ cared for/ **the happy widows**/ definitely/ {was/were} reluctant/ to work/ long shifts.

Experiment 2

Lower predictability on attractor N → Stronger AA effect?

- Results 2a: Binary categorization of predictability

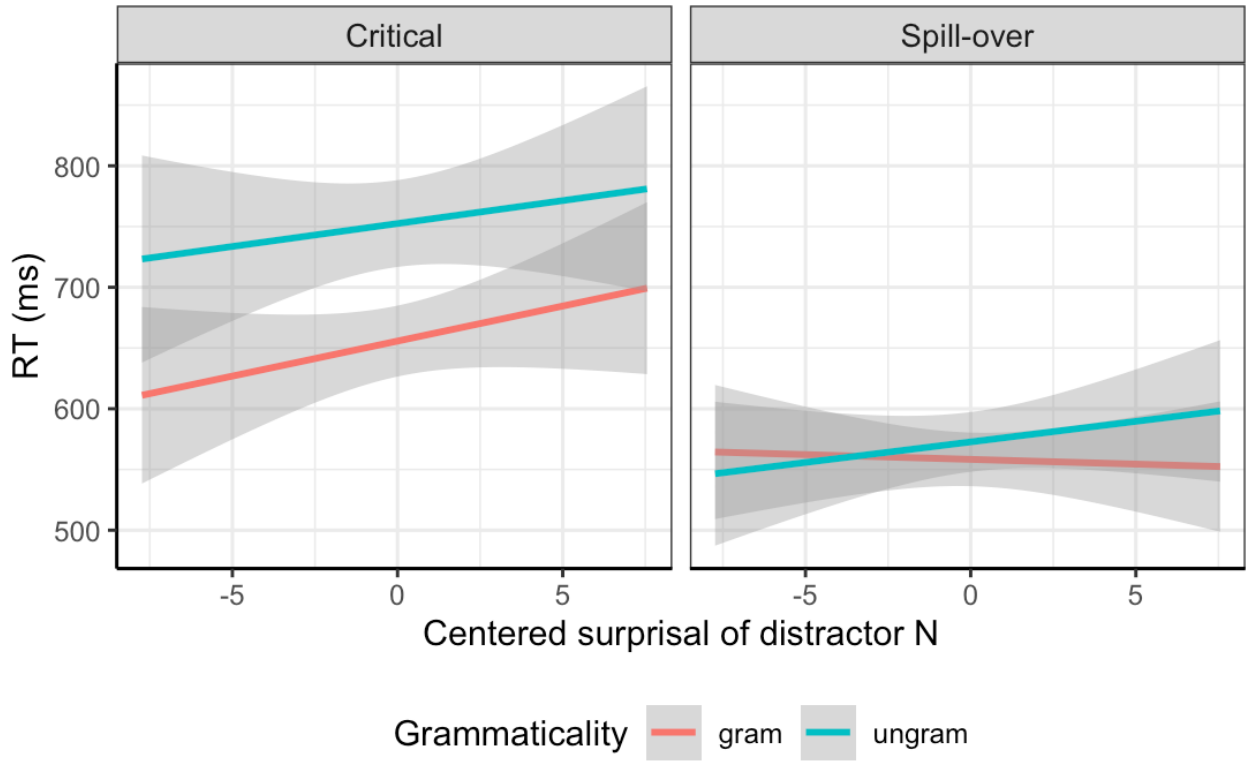


No evidence for agreement attraction effect regardless of the predictability condition

Experiment 1

Lower predictability on target N → Weaker AA effect?

- Results 2b: Surprisal of head N in subject NP generated by GPT-3



No evidence for Gram x Surprisal interaction neither in the critical nor in the spill-over region

Discussion

Memory precision in the agreement attraction effect

☑ **Lower predictability on target N → Weaker AA effect** **[Experiment 1]**

- We did observe weaker AA effect if higher surprisal on target N
- Support: Newsworthy information yields better memory representation

☒ **Lower predictability on attractor N → Stronger AA effect** **[Experiment 2]**

- The predictability/surprisal of attractor N does not modulate AA magnitude

Conclusion

- Linguistic unites with higher surprisal do yield more precise and robust memory representation
- Strategic allocation of limited memory resources to better encode unexpected but newsworthy information
- Predictability as a factor that can contribute to the variability of agreement attraction effect

References

[1] Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Second meeting of the north american chapter of the association for computational linguistics*.

[2] Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177.

[3] Rohde, H., Futrell, R., & Lucas, C. G. (2021). What's new? A comprehension bias in favor of informativity. *Cognition*, *209*, 104491.

[4] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379-423.

[5] Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and cognitive processes*, *26*(3), 376-405.

[6] Bruning, A. L., & Lewis-Peacock, J. A. (2020). Long-term memory guides resource allocation in working memory. *Scientific Reports*, *10*(1), 1-10.

[7] Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive psychology*, *23*(1), 45-93.

[8] Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of memory and language*, *61*(2), 206-237.

[9] Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*(2), 85-103.

[10] Brown, N., Bakhtin, A., Lerer, A., & Gong, Q. (2020). Combining deep reinforcement learning and search for imperfect-information games. *Advances in Neural Information Processing Systems*, 33, 17057-17069.