Is Structure Dependence Shaped for Efficient Communication?: A Case Study on Coordination

Kohei Kajikawa [1,2] Yusuke Kubota [2] Yohei Oseki [1]
[1]The University of Tokyo [2]NINJAL

CoNLL 2024@Miami

kohei-kajikawa@g.ecc.u-tokyo.ac.jp

November 15th, 2024

Efficient Communication and Structure Dependence

- ***Why*** do universal properties exist in natural languages?

Efficient Communication and Structure Dependence

- ***Why*** do universal properties exist in natural languages?
- One hypothesis:
  Universals are shaped for ***efficient communication*** [JT11; KXR18; Gib+19].

Efficient Communication and Structure Dependence

- ***Why*** do universal properties exist in natural languages?
- One hypothesis:
  Universals are shaped for ***efficient communication*** [JT11; KXR18; Gib+19].
  - The amount of information conveyed is maximized (*informative*) while the effort for usage is minimized (*simple*).
  - optimized under simplicity/informativeness trade-off.

Efficient Communication and Structure Dependence

- ***Why*** do universal properties exist in natural languages?
- One hypothesis:
  Universals are shaped for ***efficient communication*** [JT11; KXR18; Gib+19].
  - The amount of information conveyed is maximized (*informative*) while the effort for usage is minimized (*simple*).
  - optimized under simplicity/informativeness trade-off.
- To what extent can the hypothesis explain universals?

Efficient Communication and Structure Dependence

- ***Why*** do universal properties exist in natural languages?
- One hypothesis:
  Universals are shaped for ***efficient communication*** [JT11; KXR18; Gib+19].
  - The amount of information conveyed is maximized (*informative*) while the effort for usage is minimized (*simple*).
  - optimized under simplicity/informativeness trade-off.
- To what extent can the hypothesis explain universals?
  - Lexical level [i.a., KR12; PTG12; Zas+18; Mol+21; Ste21]

Efficient Communication and Structure Dependence

- ***Why*** do universal properties exist in natural languages?
- One hypothesis:
  Universals are shaped for ***efficient communication*** [JT11; KXR18; Gib+19].
  - The amount of information conveyed is maximized (*informative*) while the effort for usage is minimized (*simple*).
  - optimized under simplicity/informativeness trade-off.
- To what extent can the hypothesis explain universals?
  - Lexical level [i.a., KR12; PTG12; Zas+18; Mol+21; Ste21]
  - Syntactic level
    - Compositionality [Kir+15]
    - (Several) Greenbergian Word Order Universals [HJF20]

Efficient Communication and Structure Dependence

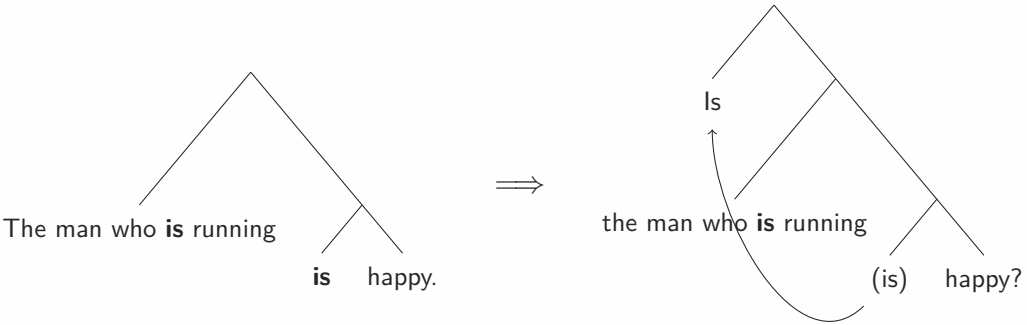
- ***Why*** do universal properties exist in natural languages?
- One hypothesis:
  Universals are shaped for ***efficient communication*** [JT11; KXR18; Gib+19].
  - The amount of information conveyed is maximized (*informative*) while the effort for usage is minimized (*simple*).
  - optimized under simplicity/informativeness trade-off.
- To what extent can the hypothesis explain universals?
  - Lexical level [i.a., KR12; PTG12; Zas+18; Mol+21; Ste21]
  - Syntactic level
    - Compositionality [Kir+15]
    - (Several) Greenbergian Word Order Universals [HJF20]
    - **Structure Dependence**? (← this work)

Structure Dependence

- Grammatical operations are applied **structurally** rather than linearly.
- In English yes-no questions,

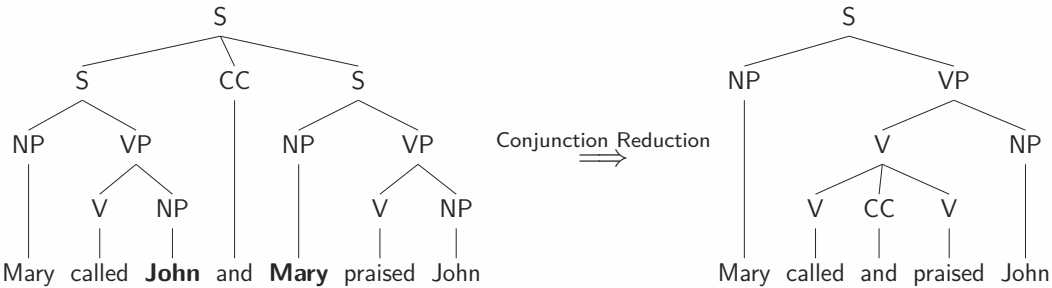
  good rule moves the auxiliary of the *main clause* to the front (structural)

  bad rule moves the *leftmost* auxiliary to the front (linear)



Structure Dependence

- **Coordinate structures** are constructed through a structure-dependent reduction operation, *conjunction reduction* [Cho57; Cho55; Ros67].
  - Which words are reduced is determined by their structural position.



Experiment: Design of 3 types of languages

1. `No-reduction` lg:
   - Mary called John and Mary praised John.
2. `Structure-reduction` lg:
   - Mary called _ and _ praised John.
3. `Linear-reduction` lg:
   - Mary called John and _ praised _.

Experiment: Design of 3 types of languages

1. `No-reduction` lg:
   - Mary called John and Mary praised John.
2. `Structure-reduction` lg:
   - Mary called _ and _ praised John.
3. `Linear-reduction` lg:
   - Mary called John and _ praised _.

- Create the corpora of them using [WC21]'s toy PCFGs.

Estimating Communicative Efficiency

- Following [HJF20], we defined *simplicity* and *informativeness* as *predictability* and *parsability*, respectively.

$$\text{predictability} := -H(\mathcal{U}) = \sum_{u \in \mathcal{U}} p(u) \log p(u) \tag{1}$$

$$\text{parsability} := -H(\mathcal{T}|\mathcal{U}) = \sum_{t \in \mathcal{T}, u \in \mathcal{U}} p(t, u) \log p(t|u) \tag{2}$$

$$\text{communicative efficiency} := \lambda \text{predictability} + (1 - \lambda)\text{parsability} \quad (\lambda \in [0, 1]) \tag{3}$$

  - Predictability is approximated with mean negative *word-by-word surprisal*.
    - represents the ease of processing on average under surprisal theory [Hal01; Lev08].
  - Parsability is approximated with mean *word-by-word logLik of parse*.
    - captures how unambiguously the underlying syntactic structure can be reconstructed.

Estimating Communicative Efficiency

- Following [HJF20], we defined *simplicity* and *informativeness* as *predictability* and *parsability*, respectively.

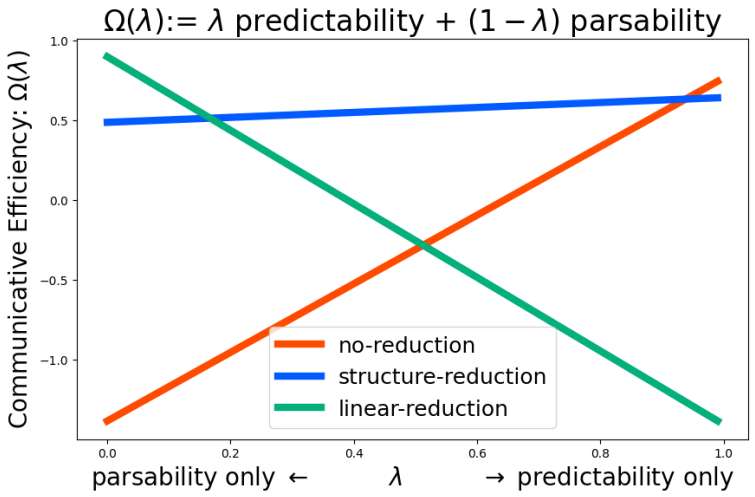
$$\text{predictability} := -H(\mathcal{U}) = \sum_{u \in \mathcal{U}} p(u) \log p(u) \tag{1}$$

$$\text{parsability} := -H(\mathcal{T}|\mathcal{U}) = \sum_{t \in \mathcal{T}, u \in \mathcal{U}} p(t, u) \log p(t|u) \tag{2}$$

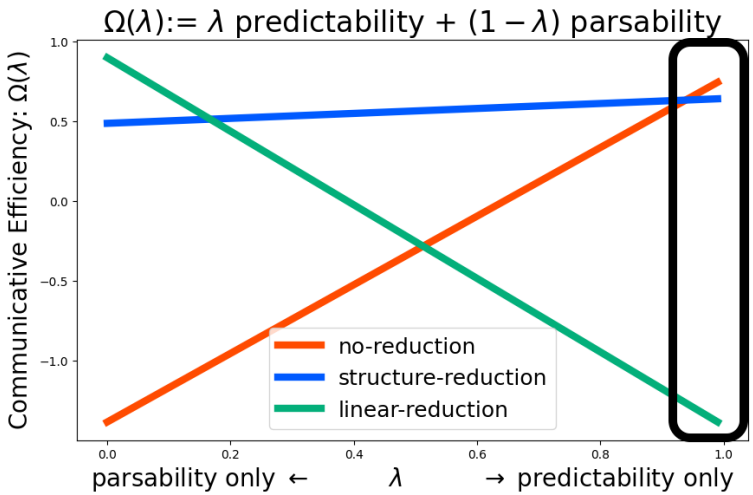
$$\text{communicative efficiency} := \lambda \text{predictability} + (1 - \lambda)\text{parsability} \quad (\lambda \in [0, 1]) \tag{3}$$

  - Predictability is approximated with mean negative *word-by-word surprisal*.
    - represents the ease of processing on average under surprisal theory [Hal01; Lev08].
  - Parsability is approximated with mean *word-by-word logLik of parse*.
    - captures how unambiguously the underlying syntactic structure can be reconstructed.
- calculated them with Reccurent Neural Network Grammars (RNNGs; [Dye+16]).

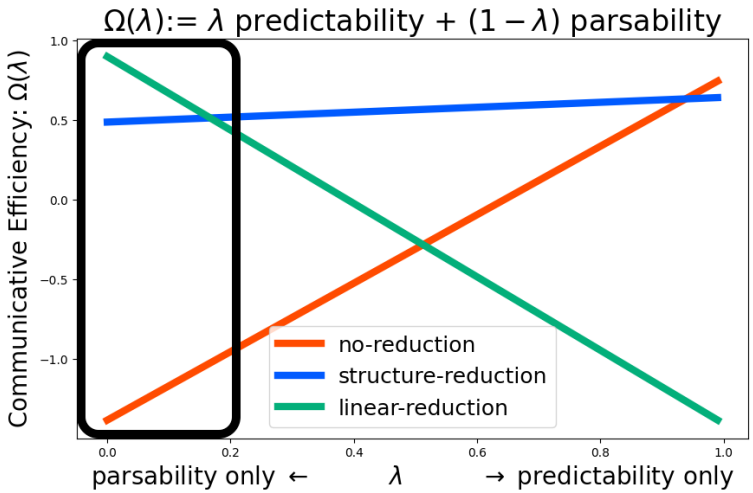
Results



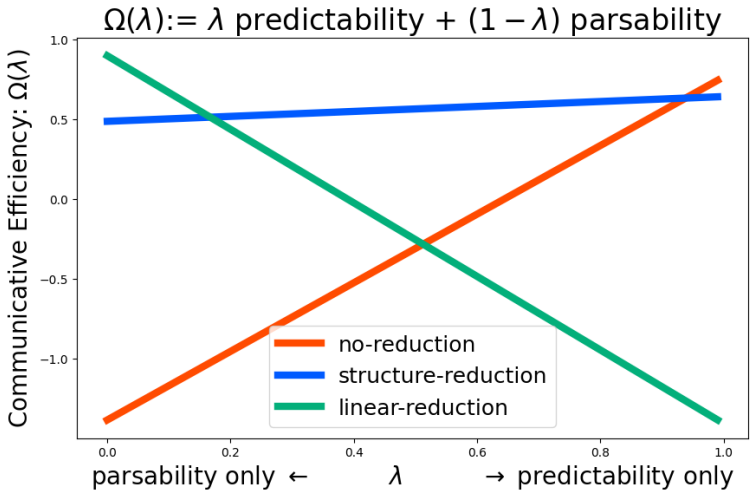
- The `structure-reduction` lgs are the most communicatively efficient under the parameter $\lambda \in [0.18, 0.93]$ for 95% CI.



- When considering only predictability (simplicity), the `no-reduction` lgs take the best score.
  - `No-reduction` lg is the simplest for local string patterns, which makes prediction easier.



- When considering only parsability (informativeness), the `linear-reduction` lgs take the best.
  - `Linear-reduction` lg has shorter overall expressions, resulting in fewer possible parses at each word position.



- Balancing the trade-off between the two, a structure-dependent reduction is the most preferred design for maximizing communicative efficiency.

Implications for Theoretical Linguistics

- A prominent view in the mainstream generative grammar:
  - natural language involves **domain-specific** predispositions and syntactic properties of language—including structure dependence—are best explained from the perspective of 'efficient *computation*' [HCF02; Cho05; Eve+15; BC16].
  - communication is considered an epiphenomenon [Cho02; HCF02].

Implications for Theoretical Linguistics

- A prominent view in the mainstream generative grammar:
  - natural language involves **domain-specific** predispositions and syntactic properties of language—including structure dependence—are best explained from the perspective of 'efficient *computation*' [HCF02; Cho05; Eve+15; BC16].
  - communication is considered an epiphenomenon [Cho02; HCF02].
- Our results suggest that at least some structure-dependent properties present in natural language (such as coordination) can be explained from the perspective of **domain-general** efficient *communication*.
  - This aligns with the existing body of efficient communication research [Gib+19; FPG24].

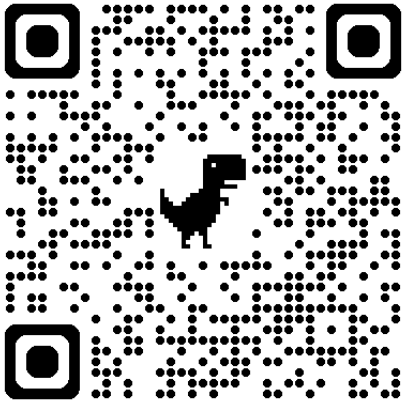
Conclusion

- We investigated whether **structure dependence** reflects the optimization for **efficient communication**.
  - focusing on coordinate structures.
- The experiment results suggest that the structure-dependent properties can be reduced to the functional perspective of efficient communication.

Thank you!

Title: Is Structure Dependence Shaped for Efficient Communication?: A Case Study on Coordination

Authors: Kohei Kajikawa, Yusuke Kubota, and Yohei Oseki



Link to Paper

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers 21K00541 and 24H00087, JST PRESTO Grant Number JPMJPR21C2, and the NINJAL collaborative research project 'Toward a Computationally-Informed Theoretical Linguistics'.

[BC16] Robert C. Berwick and Noam Chomsky. *Why Only Us: Language and Evolution*. Cambridge, MA: The MIT Press, 2016. URL: https://mitpress.mit.edu/9780262533492/why-only-us/.

[Cho02] Noam Chomsky. "An interview on minimalism". In: *On Nature and Language*. Ed. by Adriana Belletti and Luigi Rizzi. Cambridge University Press, 2002, pp. 92–161. DOI: 10.1017/CBO9780511613876.005.

[Cho05] Noam Chomsky. "Three Factors in Language Design". In: *Linguistic Inquiry* 36.1 (Jan. 2005), pp. 1–22. ISSN: 0024-3892. DOI: 10.1162/0024389052993655.

[Cho55] Noam Chomsky. *The Logical Structure of Linguistic Theory*. Springer New York, NY, 1975 (=1955). URL: https://link.springer.com/book/9780306307607.

[Cho57] Noam Chomsky. *Syntactic Structures*. Mouton, 1957. DOI: 10.1515/9783112316009.

[Dye+16] Chris Dyer et al. "Recurrent Neural Network Grammars". In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego, California: Association for Computational Linguistics, June 2016, pp. 199–209. DOI: 10.18653/v1/N16-1024.

[Eve+15] Martin B.H. Everaert et al. "Structures, Not Strings: Linguistics as Part of the Cognitive Sciences". In: *Trends in Cognitive Sciences* 19 (12 2015), pp. 729–743. DOI: 10.1016/j.tics.2015.09.008.

[FPG24] Evelina Fedorenko, Steven T. Piantadosi, and Edward A. F. Gibson. "Language is primarily a tool for communication rather than thought". In: *Nature* 630 (8017 2024), pp. 575–586. DOI: 10.1038/s41586-024-07522-w.

[Gib+19] Edward Gibson et al. "How Efficiency Shapes Human Language". In: *Trends in Cognitive Sciences* 23.5 (2019), pp. 389–407. ISSN: 1364-6613. DOI: 10.1016/j.tics.2019.02.003.

[Hal01] John Hale. "A Probabilistic Earley Parser as a Psycholinguistic Model". In: *Second Meeting of the North American Chapter of the Association for Computational Linguistics*. 2001. URL: https://aclanthology.org/N01-1021.

[HCF02] Marc D. Hauser, Noam Chomsky, and W. Tecumseh Fitch. "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" In: *Science* 298.5598 (2002), pp. 1569–1579. DOI: 10.1126/science.298.5598.1569.

[HJF20] Michael Hahn, Dan Jurafsky, and Richard Futrell. "Universals of word order reflect optimization of grammars for efficient communication". In: *Proceedings of the National Academy of Sciences* 117.5 (2020), pp. 2347–2353. DOI: 10.1073/pnas.1910923117.

[JT11] T. Florian Jaeger and Harry Tily. "On language utility: processing complexity and communicative efficiency". In: *WIREs Cognitive Science* 2.3 (2011), pp. 323–335. DOI: 10.1002/wcs.126.

[Kir+15] Simon Kirby et al. "Compression and communication in the cultural evolution of linguistic structure". In: *Cognition* 141 (2015), pp. 87–102. ISSN: 0010-0277. DOI: 10.1016/j.cognition.2015.03.016.

[KR12] Charles Kemp and Terry Regier. "Kinship Categories Across Languages Reflect General Communicative Principles". In: *Science* 336.6084 (2012), pp. 1049–1054. DOI: 10.1126/science.1218811.

[KXR18] Charles Kemp, Yang Xu, and Terry Regier. "Semantic Typology and Efficient Communication". In: *Annual Review of Linguistics* 4.1 (2018), pp. 109–128. DOI: 10.1146/annurev-linguistics-011817-045406.

[Lev08] Roger Levy. "Expectation-based syntactic comprehension". In: *Cognition* 106.3 (2008), pp. 1126–1177. ISSN: 0010–0277. DOI: 10.1016/j.cognition.2007.05.006.

[Mol+21] Francis Mollica et al. "The forms and meanings of grammatical markers support efficient communication". In: *Proceedings of the National Academy of Sciences* 118.49 (2021), e2025993118. DOI: 10.1073/pnas.2025993118.

[PTG12] Steven T. Piantadosi, Harry Tily, and Edward Gibson. "The communicative function of ambiguity in language". In: *Cognition* 122.3 (2012), pp. 280–291. ISSN: 0010-0277. DOI: https://doi.org/10.1016/j.cognition.2011.10.004. URL: https://www.sciencedirect.com/science/article/pii/S0010027711002496.

[Ros67] John R. Ross. “Constraints on variables in syntax”. PhD thesis. MIT, 1967.

[Ste21] Shane Steinert-Threlkeld. “Quantifiers in Natural Language: Efficient Communication and Degrees of Semantic Universals”. In: *Entropy* 23.10 (2021). ISSN: 1099-4300. DOI: 10.3390/e23101335.

[WC21] Jennifer C. White and Ryan Cotterell. “Examining the Inductive Bias of Neural Language Models with Artificial Languages”. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Ed. by Chengqing Zong et al. Online: Association for Computational Linguistics, Aug. 2021, pp. 454–463. DOI: 10.18653/v1/2021.acl-long.38.

[Zas+18] Noga Zaslavsky et al. “Efficient compression in color naming and its evolution”. In: *Proceedings of the National Academy of Sciences* 115.31 (2018), pp. 7937–7942. DOI: 10.1073/pnas.1800521115.