

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

Kohei Kajikawa

The University of Tokyo / NINJAL

@NLP Colloquium

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

January 29th, 2025

自己紹介

- ▶ 梶川 康平 (かじかわ こうへい)
 - ▶ 東京大学大学院 大関研究室 修士2年
- ▶ 研究：計算心理言語学
 - ▶ 人間は文理解の際に文法知識をどのように使っているのか ([Kajikawa et al., 2024b](#))
 - ▶ CCG による文理解のモデル化
 - ▶ そうした文法知識がなぜ今ある形になっているのか ([Kajikawa et al., 2024a](#))



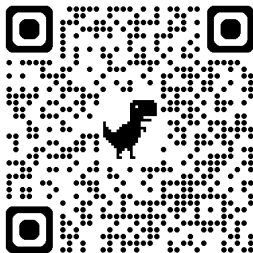
↑Web ページ

CoNLL 2024@Miami

- ▶ 窪田悠介先生、大関洋平先生との研究 ([Kajikawa et al., 2024a](#)) を話します



- ▶ 論文リンク



認知負荷の最適化戦略としての自由語順と項省略

梶川康平^{1,2} 磯野真之介^{1,2} 窪田悠介² 大関洋平¹

¹ 東京大学 ² 国立国語研究所

{kohei-kajikawa, isono-shinnosuke, osekai}@g.ecc.u-tokyo.ac.jp kubota@ninjal.ac.jp

概要

自由語順と項省略という統語現象には相関関係がみられる。つまり、言語ごとに、両方の現象が観察されるか、どちらも観察されないか、という傾向がある。では、なぜそのような文法相関が存在するのだろうか。本研究では、自由語順と項省略がみられる日本語に注目し、処理の効率性がこれらの特徴を形作っているのか検証する。具体的には、日本語のコーパスから、自由語順、項省略、またはその両方



図1 本研究の概要。自由語順と項省略という相関のある統語現象について、両方同時に存在することで処理の効率性が高まるのか検証する。

(c) 「処理のしやすさ」と文法の関係

と、そのほか共著研究...

CCG による日本語脳波データのモデリング

磯野 真之介^{1,3,4*} 梶川 康平^{1,3*} 杉本侑嗣^{2*} 浅原正幸³ 大関洋平¹

¹ 東京大学 ² 大阪大学

³ 国立国語研究所 ⁴ 日本学術振興会特別研究員 (DC2)
{isono-shinnosuke, kohei-kajikawa, osekai}@g.ecc.u-tokyo.ac.jp

sugimoto.yushi.hmt@osaka-u.ac.jp

masayu-a@ninjal.ac.jp

概要

計算心理言語学では、人間の文理解において、どのような表象がどのように構築されているのかが探究されてきた。この問いに取り組むのに、Combinatory Categorical Grammar (CCG) が有望な文法理論であることが、日本語や英語の読み時間や脳血

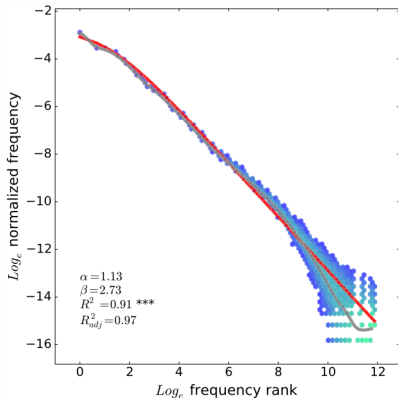
脳波データを予測でき、できるとすればどの脳波成分に対応するかも明らかでない。

そこで、本研究では日本語を対象として、新たな脳波データセットを構築した上で、CCG に基づく複数の指標を用いて脳波データを分析する。分析にあたっては、他の文法 (文脈自由文法および依存文法) に基づく指標や、Transformer ベースの言語モデ

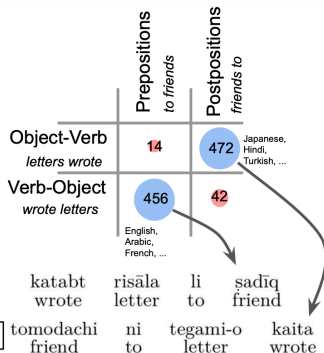
(d) CCG による文処理研究

言語の普遍性

- ▶ 自然言語には数多くの普遍的な性質 (universals/strong statistical tendencies) がある (Zipf, 1949; Greenberg, 1963; Dryer, 1992; Christiansen et al., 2009)



(e) 単語頻度の関係。Piantadosi (2014) より



(f) 語順の普遍性。Hahn et al. (2020) より

効率的なコミュニケーション

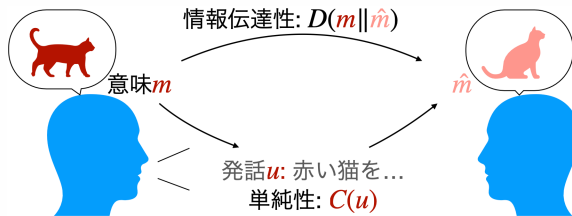
- ▶ 自然言語にはなぜ普遍性が存在するのか？

効率的なコミュニケーション

- ▶ 自然言語にはなぜ普遍性が存在するのか？
- ▶ 一つの仮説として：
普遍性は、効率的なコミュニケーションの実現の結果である ([Jaeger and Tily, 2011](#);
[Kemp et al., 2018](#); [Gibson et al., 2019](#))

効率的なコミュニケーション

- ▶ 自然言語にはなぜ普遍性が存在するのか？
- ▶ 一つの仮説として：
普遍性は、効率的なコミュニケーションの実現の結果である (Jaeger and Tily, 2011; Kemp et al., 2018; Gibson et al., 2019)
 - ▶ 伝わる意図・情報が最大化されている (*informative*) 一方、使用のコストが最小化されている (*simple*) 状況
 - ▶ 単純性 (simplicity) と情報伝達性 (informativeness) のトレードオフのものの最適化



効率的なコミュニケーション

- ▶ では、この仮説で普遍性をどこまで説明可能？

効率的なコミュニケーション

- ▶ では、この仮説で普遍性をどこまで説明可能？
 - ▶ 語彙 (wordform/semantic categorization) (i.a., [Kemp and Regier, 2012](#); [Piantadosi et al., 2012](#); [Zaslavsky et al., 2018](#); [Mollica et al., 2021](#); [Steinert-Threlkeld, 2021](#))

効率的なコミュニケーション

▶ では、この仮説で普遍性をどこまで説明可能？

- ▶ 語彙 (wordform/semantic categorization) (i.a., [Kemp and Regier, 2012](#); [Piantadosi et al., 2012](#); [Zaslavsky et al., 2018](#); [Mollica et al., 2021](#); [Steinert-Threlkeld, 2021](#))
- ▶ 文法
 - ▶ 構成性 ([Kirby et al., 2015](#); [Futrell and Hahn, 2024](#))
 - ▶ (一部の) 語順普遍 ([Hahn et al., 2020](#))

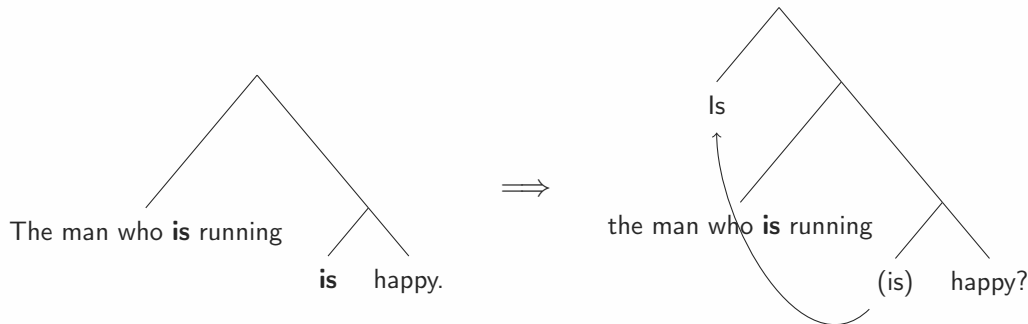
効率的なコミュニケーション

▶ では、この仮説で普遍性をどこまで説明可能？

- ▶ 語彙 (wordform/semantic categorization) (i.a., [Kemp and Regier, 2012](#); [Piantadosi et al., 2012](#); [Zaslavsky et al., 2018](#); [Mollica et al., 2021](#); [Steinert-Threlkeld, 2021](#))
- ▶ 文法
 - ▶ 構成性 ([Kirby et al., 2015](#); [Futrell and Hahn, 2024](#))
 - ▶ (一部の) 語順普遍 ([Hahn et al., 2020](#))
 - ▶ 構造依存性 ([Kajikawa et al., 2024a](#)) (← 今日の話)

構造依存性

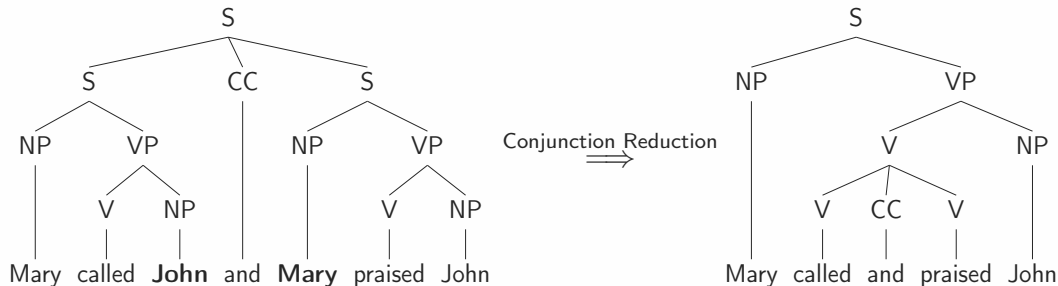
- ▶ 文法操作は線形ではなく構造依存で適用される
- ▶ たとえば、英語の yes-no 疑問文だと：
 - ✕ 一番左側にある助動詞を文頭にもっていく（線形）
 - 主節の助動詞を文頭にもっていく（構造依存）



構造依存性

- ▶ 等位接続構造もまた、構造依存の縮約操作によって構成される (Chomsky, 1957, 1955; Ross, 1967)

- ▶ どの単語が縮約されるかはその構造上の位置で決まる



実験：3種類の言語のデザイン

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 _.

実験：3種類の言語のデザイン

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 _.

	縮約あり	縮約なし
構造依存	structure-reduction	no-reduction
線形	linear-reduction	

実験：3種類の言語のデザイン

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 _.

	縮約あり	縮約なし
構造依存	structure-reduction	no-reduction
線形	linear-reduction	

- ▶ White and Cotterell (2021) の人工 PCFGs でコーパスを作成

コミュニケーションの効率性の推定

- ▶ Hahn et al. (2020) に従い、単純性と情報伝達性をそれぞれ *predictability* と *parsability* で定義

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

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

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

- ▶ Predictability は単語ごとの平均サプライザルで近似
 - ▶ サプライザル理論 (Hale, 2001; Levy, 2008) における平均的な処理のしやすさ
- ▶ Parsability は単語ごとの構文木の対数尤度の平均で近似
 - ▶ どれだけ曖昧性なく、背後の統語構造を復元できるか

コミュニケーションの効率性の推定

- ▶ Hahn et al. (2020) に従い、単純性と情報伝達性をそれぞれ *predictability* と *parsability* で定義

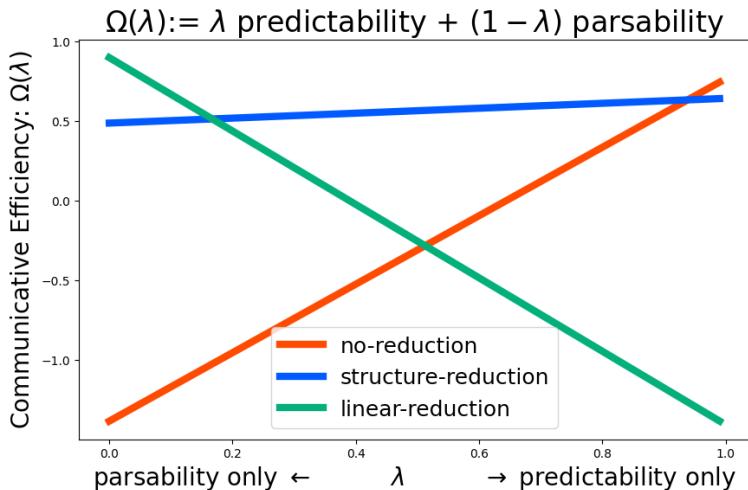
$$\text{predictability} := -H(\mathcal{U}) = \sum_{u \in \mathcal{U}} p(u) \log p(u) \quad (1)$$

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

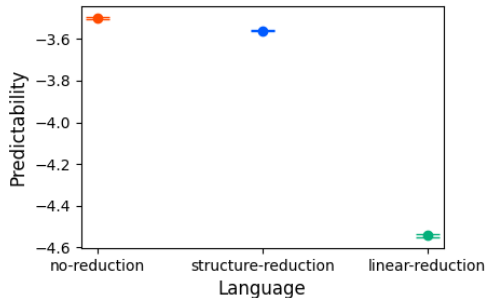
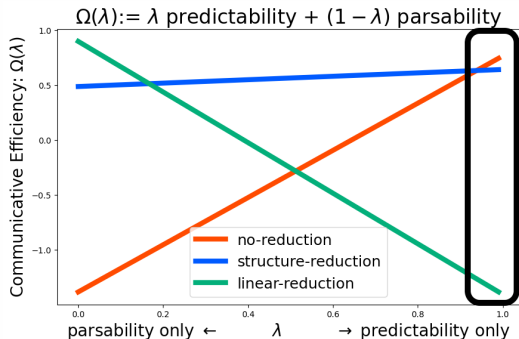
$$\text{communicative efficiency} := \lambda \cdot \text{predictability} + (1 - \lambda) \cdot \text{parsability} \quad (\lambda \in [0, 1]) \quad (3)$$

- ▶ Predictability は単語ごとの平均サプライザルで近似
 - ▶ サプライザル理論 (Hale, 2001; Levy, 2008) における平均的な処理のしやすさ
- ▶ Parsability は単語ごとの構文木の対数尤度の平均で近似
 - ▶ どれだけ曖昧性なく、背後の統語構造を復元できるか
- ▶ Recurrent Neural Network Grammars (RNNGs; Dyer et al., 2016) で推定

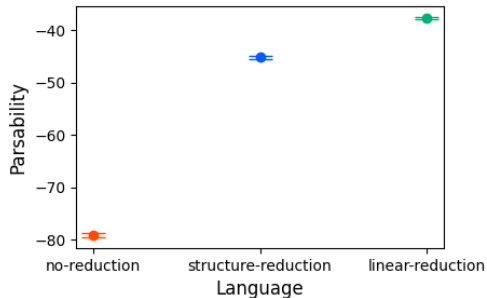
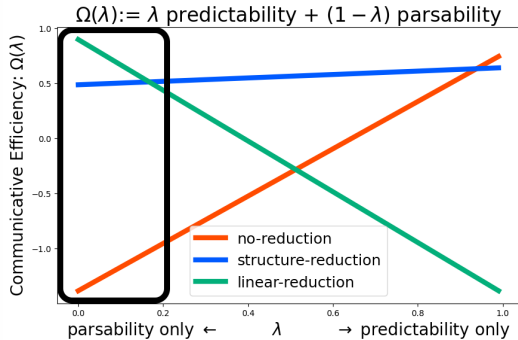
結果



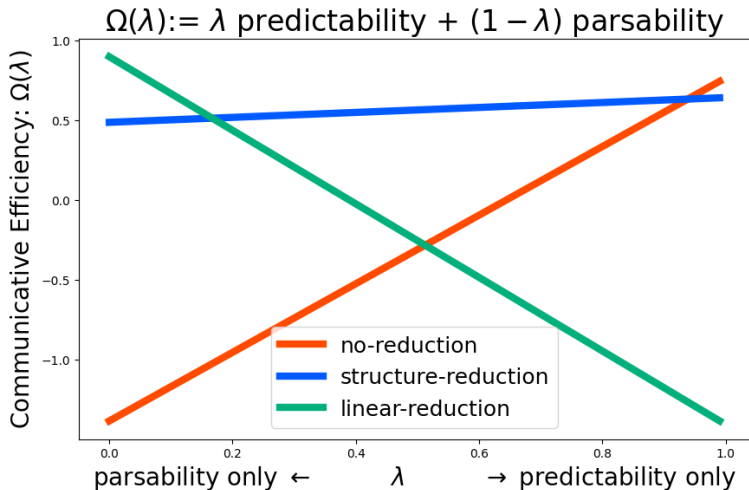
- ▶ Structure-reduction lgs が、 $\lambda \in [0.18, 0.93]$ のもとで最もコミュニケーション上効率的



- ▶ Predictability (単純性) のみ考慮すると、no-reduction lgs が最大値を取る
 - ▶ No-reduction lg は局所的な文字列のパターンが最も単純であり、それにより次単語予測がしやすい



- ▶ Parsability (情報伝達性) のみ考慮すると、linear-reduction lgs が最大値を取る
 - ▶ Linear-reduction lg は全体的に短い傾向にあり、各単語時点での可能な構文木が少ない傾向



- ▶ 両者のバランスをとり、構造依存な縮約がコミュニケーションの効率性を上げるのに最も好まれるデザインである

理論言語学への示唆

- ▶ 主流の生成文法における主要な見解：
 - ▶ 自然言語は、**領域固有** (domain-specific) の言語知識を伴い、言語の統語的性質（構造依存性など）は、「効率的な**計算**」という観点から説明される ([Hauser et al., 2002](#); [Chomsky, 2005](#); [Everaert et al., 2015](#); [Berwick and Chomsky, 2016](#))
 - ▶ コミュニケーションは付随的な現象 (epiphenomena) ([Chomsky, 2002](#); [Hauser et al., 2002](#))

理論言語学への示唆

- ▶ 主流の生成文法における主要な見解：
 - ▶ 自然言語は、**領域固有** (domain-specific) の言語知識を伴い、言語の統語的性質 (構造依存性など) は、「効率的な**計算**」という観点から説明される (Hauser et al., 2002; Chomsky, 2005; Everaert et al., 2015; Berwick and Chomsky, 2016)
 - ▶ コミュニケーションは付随的な現象 (epiphenomena) (Chomsky, 2002; Hauser et al., 2002)
- ▶ 我々の結果は、自然言語に見られる構造依存性が、**領域一般** (domain-general) の効率的な**コミュニケーション**という観点から説明可能であることを示唆
 - ▶ 効率的なコミュニケーションに関する既存研究と一致 (Gibson et al., 2019; Fedorenko et al., 2024)

「正直な」 limitations

1. Communicative efficiency の目的関数の改善

「正直な」 limitations

1. Communicative efficiency の目的関数の改善

- ▶ 単純性では、記憶制約 ([Gibson, 1998](#); [Lewis and Vasishth, 2005](#); [Isono, 2024](#)) も考慮したい
 - ▶ 予測と記憶にはトレードオフがある ([Futrell et al., 2020](#); [Hahn et al., 2021](#))
- ▶ 情報伝達性を parsability で評価したが、これは心理的に妥当か？

「正直な」 limitations

1. Communicative efficiency の目的関数の改善

- ▶ 単純性では、記憶制約 (Gibson, 1998; Lewis and Vasishth, 2005; Isono, 2024) も考慮したい
 - ▶ 予測と記憶にはトレードオフがある (Futrell et al., 2020; Hahn et al., 2021)
- ▶ 情報伝達性を parsability で評価したが、これは心理的に妥当か？

2. 学習可能性 (learnability) との関係？

「正直な」 limitations

1. Communicative efficiency の目的関数の改善

- ▶ 単純性では、記憶制約 (Gibson, 1998; Lewis and Vasishth, 2005; Isono, 2024) も考慮したい
 - ▶ 予測と記憶にはトレードオフがある (Futrell et al., 2020; Hahn et al., 2021)
- ▶ 情報伝達性を parsability で評価したが、これは心理的に妥当か？

2. 学習可能性 (learnability) との関係？

3. 自然言語データの使用、扱う現象の拡張

結論

- ▶ 構造依存性が効率的なコミュニケーションの最適化を反映しているか？
 - ▶ 等位接続構造に注目し検証
- ▶ 結果、構造依存性が効率的なコミュニケーションという機能的観点から説明可能であることを示唆

Reference I

- Berwick, R. C. and Chomsky, N. (2016). *Why Only Us: Language and Evolution*. The MIT Press, Cambridge, MA.
- Chomsky, N. (1957). *Syntactic Structures*. Mouton.
- Chomsky, N. (1975/1955). *The Logical Structure of Linguistic Theory*. Springer New York, NY.
- Chomsky, N. (2002). An interview on minimalism. In Belletti, A. and Rizzi, L., editors, *On Nature and Language*, pages 92–161. Cambridge University Press.
- Chomsky, N. (2005). Three Factors in Language Design. *Linguistic Inquiry*, 36(1):1–22.
- Christiansen, M. H., Collins, C., and Edelman, S. (2009). *Language Universals*. Oxford University Press.
- Dryer, M. S. (1992). The Greenbergian word order correlations. *Language*, 68(1):81–138.
- Dyer, C., Kuncoro, A., Ballesteros, M., and Smith, N. A. (2016). Recurrent neural network grammars. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 199–209, San Diego, California. Association for Computational Linguistics.
- Everaert, M. B., Huybregts, M. A., Chomsky, N., Berwick, R. C., and Bolhuis, J. J. (2015). Structures, not strings: Linguistics as part of the cognitive sciences. *Trends in Cognitive Sciences*, 19:729–743.
- Fedorenko, E., Piantadosi, S. T., and Gibson, E. A. F. (2024). Language is primarily a tool for communication rather than thought. *Nature*, 630:575–586.
- Futrell, R., Gibson, E., and Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science*, 44(3):e12814.
- Futrell, R. and Hahn, M. (2024). Linguistic structure from a bottleneck on sequential information processing.

Reference II

- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68(1):1–76.
- Gibson, E., Futrell, R., Piantadosi, S. T., Dautriche, I., Mahowald, K., Bergen, L., and Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, 23(5):389–407.
- Greenberg, J. H. (1963). *Universals of Language*. MIT Press.
- Hahn, M., Degen, J., and Futrell, R. (2021). Modeling word and morpheme order in natural language as an efficient tradeoff of memory and surprisal. *Psychological Review*, 128:726–756.
- Hahn, M., Jurafsky, D., and Futrell, R. (2020). Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences*, 117(5):2347–2353.
- 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*.
- Hauser, M., Chomsky, N., and Fitch, W. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598):1569–1579.
- Isono, S. (2024). Category locality theory: A unified account of locality effects in sentence comprehension. *Cognition*, 247:105766.
- Jaeger, T. F. and Tily, H. (2011). On language ‘utility’: processing complexity and communicative efficiency. *WIREs Cognitive Science*, 2(3):323–335.
- Kajikawa, K., Kubota, Y., and Oseki, Y. (2024a). Is structure dependence shaped for efficient communication?: A case study on coordination. In Barak, L. and Alikhani, M., editors, *Proceedings of the 28th Conference on Computational Natural Language Learning*, pages 291–302, Miami, FL, USA. Association for Computational Linguistics.

Reference III

- Kajikawa, K., Yoshida, R., and Oseki, Y. (2024b). Dissociating Syntactic Operations via Composition Count. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 46, pages 297–305, Rotterdam, the Netherlands.
- Kemp, C. and Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, 336(6084):1049–1054.
- Kemp, C., Xu, Y., and Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics*, 4(1):109–128.
- Kirby, S., Tamariz, M., Cornish, H., and Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.
- Lewis, R. L. and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3):375–419.
- Mollica, F., Bacon, G., Zaslavsky, N., Xu, Y., Regier, T., and Kemp, C. (2021). The forms and meanings of grammatical markers support efficient communication. *Proceedings of the National Academy of Sciences*, 118(49):e2025993118.
- Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, 21(5):1112–1130.
- Piantadosi, S. T., Tily, H., and Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122(3):280–291.

Reference IV

- Ross, J. R. (1967). *Constraints on variables in syntax*. PhD thesis, MIT.
- Steinert-Threlkeld, S. (2021). Quantifiers in natural language: Efficient communication and degrees of semantic universals. *Entropy*, 23(10).
- White, J. C. and Cotterell, R. (2021). Examining the Inductive Bias of Neural Language Models with Artificial Languages. In Zong, C., Xia, F., Li, W., and Navigli, R., editors, *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)*, pages 454–463, Online. Association for Computational Linguistics.
- Zaslavsky, N., Kemp, C., Regier, T., and Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, 115(31):7937–7942.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley.