

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¹¹ 東京大学 ² 国立国語研究所

{kohei-kajikawa, isono-shinnosuke, oseki}@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を検証する。具体的には、日本語のコーパスから、自由語順、項省略、またはその両方がないバージョンを作成し、現実のコーパスと比較する。その結果、自由語順かつ項省略が存在する言語は、記憶と予測の負荷を抑制する点で有利だと示された。

1 はじめに

世界に7,000以上あるとされる自然言語には、非常に多様な性質がある[1]。一方で、こうした多様性とは裏腹に、言語間で共通して見られる普遍的な性質も確かに存在し、それらは何らかの原理に還元できる可能性がある。では、それぞれの言語は、なぜいまあるような形をとっているのだろうか。

一つの仮説として、自然言語は効率的なコミュニケーションを実現するよう形作られている、というものがある[2, 3, 4, 5]。コミュニケーションが効率的であるとは、人間の認知制約のもと、伝わる意図・情報が最大化されている一方、産出や理解といった使用のコストが最小化されている状態のことを指す。コミュニケーションにおける効率性が、言語そのものを形作っているという考えは、言語学では古くより提案されているが[6, 7, 8, 9]、近年、心理言語学における言語使用コストの定量化理論や、大規模な言語資源の整備を背景として、この仮説を計算機上で検証する方法論が確立されつつある。これまで、特に語彙に関しては活発に研究が行われており(e.g., [3, 10, 11])、さらに統語レベルでも、構成性[12]や語順の普遍性[13]、そして等位接続にお

自由語順 (かき混ぜ)

花子が 太郎を ほめた (主語—目的語—動詞)

太郎を 花子が ほめた (目的語—主語—動詞)

項省略

花子が 太郎を ほめた (主語の省略)

花子が 太郎を ほめた (目的語の省略)

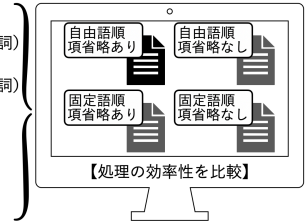


図1 本研究の概要。自由語順と項省略という相関のある統語現象について、両方同時に存在することで処理の効率性が高まるのか検証する。

る構造依存性[14]など、複数の統語的普遍性について、検証が進みつつある。

では、統語現象の相関関係についても、同様にコミュニケーションの効率性の観点から説明できるのであろうか。例えば、自由語順と項省略は、複数の言語で共通に観察される一方、どちらの現象も見られない言語も存在するなど、興味深い相関関係がある[15, 16, 17, 18]。こうした文法相関について、これまで特に生成文法の文脈では、どのような一般化が可能か理論的な研究がなされてきたが、さらに、そもそもなぜそのような相関関係が自然言語に存在するのか、コミュニケーションの効率性の観点からコーパスを用いて実証的に検証する。

本研究では、自由語順と項省略の両方が観察される日本語を対象として、これらの現象が同時に存在することにより言語コミュニケーション上の効率性が高まるのか検証する。具体的には、日本語のUDコーパス(UD_Japanese-BCCWJ[19])および、日本語項省略判断(JAOJ[20])を用いて、自由語順、項省略、またはその両方がないバージョンを作成する。そして、それぞれのコミュニケーションの効率性を記憶と予測における処理負荷の観点から比較する(図1)。結果として、自由語順と項省略が両方存在することは、記憶と予測のトレードオフのもとで処理負荷を小さくするのに寄与し、特に記憶負荷の削減に重きを置いた最適化の結果であることを示す。

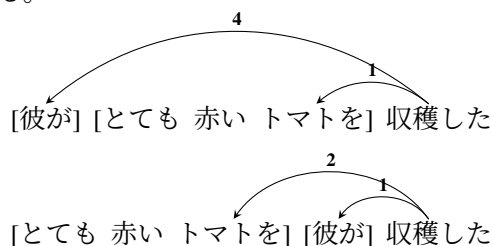
2 記憶と予測のトレードオフ

効率的なコミュニケーション仮説では、しばしば単純性 (simplicity) と情報伝達性 (informativeness) のトレードオフが想定される [3, 4, 10, 11, 12, 13, 14]。

ここで単純性とは、使用 (産出・理解) に関する処理負荷の小ささを指し、情報伝達性とは、話し手の意図が聞き手に正しく推定される程度を指す。本研究では、このうち、前者の単純性の内部における記憶と予測のトレードオフについて議論する。

これまでの心理言語学研究において、文の処理負荷は**記憶**と**予測**の2つが要因であるとされている [21, 22, 23, 24, 25]。

記憶に関わる処理負荷 Dependency Locality Theory (DLT) [21] をはじめとした文理解の理論では、作業記憶での手続きで負荷が生じると考える。文の入力に対して、その背後にある統語・意味構造を構築するのに、関連する単語間の距離が遠ければ遠いほど、その分単語を作業記憶に長く留めておかねばならず、いざ必要となるときに作業記憶から取り出すのに困難が生じると考えられる。DLTでは、こうした直感を依存文法を用いて説明する。具体的には、依存関係にある単語が互いに離れているほど処理負荷が高まるとする。こうした記憶に関する処理負荷は、語順選択における選好も説明することができ、たとえば、日本語では、動詞の前に複数の名詞句を並べるとき、短い句より先に長い句が置かれる語順 (“long-before-short” 語順) が好まれるが [8, 26, 27]、これはまさに依存関係の距離の観点で説明することができる。



上図に示すように、“long-before-short” 語順の方が、日本語のような動詞後置言語では依存関係の距離が短くなる。

予測に関わる処理負荷 記憶とは別に、予測のしにくさもまた文の処理負荷に関わる。サプライザル理論 [23, 24] では、文脈 C のもとでの単語 w の確率の負の対数 $-\log P(w | C)$ (サプライザル) を単語 w の処理負荷とする。これは読み時間と比例すること

が経験的に知られている [28, 29]。また、日本語の項省略は、後続動詞のサプライザルが大きくならないよう行われているという観察もある [30]。

そして、記憶と予測に関する処理負荷については、トレードオフ関係がある [25, 31]。文脈 C が大きければ大きいほど、それを踏まえた単語 w の予測は行いやすくなる。しかしながら、同時に、文脈の情報を維持しておく記憶のコストが大きくなる。反対に、維持しておく先行文脈を減らすと記憶の負荷は小さくなるが、単語の予測はその分難しくなる。

自然言語の語順は、この記憶と予測のトレードオフのもとで、効率的なように形作られていることが示されている [31, 32, 33]。また、自由語順と項省略は、記憶と予測の処理負荷に大きな影響を与える。これを踏まえ、本研究は、自由語順と項省略の相関関係が記憶と予測のトレードオフにより説明できるか検証する。

3 実験

日本語において、自由語順と項省略が両方同時に存在することで処理の効率性が上がるのか検証するため、自由語順と項省略の存在をコントロールしたコーパスを作成し、それぞれを記憶と予測の処理負荷の観点から評価する。

3.1 データ

日本語の依存構造アノテーション付きコーパスである UD_Japanese-BCCWJ [19] (v2.10, 長単位分割) と、項省略において、その省略されている項と位置がアノテーションされている Japanese Argument Omission Judgment (JAOJ) [20] を用いる。それぞれのデータは、ともに現代日本語書き言葉均衡コーパス (BCCWJ) [34] の書籍 (PB) ドメインの文章を用いており、重ね合わせ可能である。これらを統合することで、(i) 依存構造がアノテーションされていてかつ、(ii) どこでどのような項が省略されているのかの情報が付与されているコーパスを構成した¹⁾。交差依存 (non-projective) の文 (42 文) を取り除き、あわせて 3,428 文を分析対象とした。全体で 67,088 の単語 (国語研長単位) で構成されており、それに加え、2,895 単語分の項が省略されている。

1) JAOJ では、どのような表現が省略されているのかの情報は付与されているが、その表現の単語区切りに関する情報は付与されていないので、格を取り除いた上で、全く同じ表現 UD_Japanese-BCCWJ で探し、その長単位の分割に合わせて分割を行う、ということを行なった。

3.2 自由語順・項省略のコントロール

動詞からの依存関係が nsubj でラベル付けされている名詞句を主語と、obj でラベル付けされている名詞句を目的語として扱う。同一文中の主語と目的語の語順を調整することで、以下の4種類の言語を作った：

- Id** 元の語順のまま (identical) の言語
- Subj>Obj** 常に主語が目的語より前になるよう並べ替えた言語
- Obj>Subj** 常に目的語が主語より前になるよう並べ替えた言語
- Swap** もとが主語・目的語の順ならば目的語・主語の順に、もとが目的語・主語の順ならば主語・目的語の順に並び替えた言語

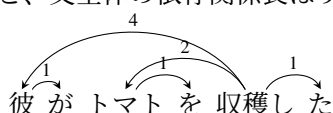
そして、JAOJ をもとに、元の文に省略された項を補うことで「項省略なし」の言語を作成する。一方、オリジナルのまま省略を維持する場合は「項省略あり」とする。つまり、語順4パターンと項省略あり／なしの2パターンの計8種類の言語を作成した。

	項省略あり	項省略なし
自由語順あり	Id	Id _{no-omission}
	Swap	Swap _{no-omission}
自由語順なし	Subj>Obj	Subj>Obj _{no-omission}
	Obj>Subj	Obj>Subj _{no-omission}

また、ベースラインとして、[35] の方法論を用いて、語順規則をランダムに再構築した文法を500種類設計し、それぞれの項省略あり／なしのあわせて1,000の言語を作った。

3.3 記憶と予測の処理負荷

依存関係長 記憶に関する処理負荷は、依存文法における長単位間の依存関係長 (dependency length) を用いて推定する。依存関係長は、依存関係間にある長単位の数+1で数える。たとえば、以下のような依存構造だと、文全体の依存関係長は9である：



各コーパスを、依存関係長を長単位数で正規化した平均依存関係長、すなわち、記憶に関する長単位ごとの平均的な処理負荷で評価する。

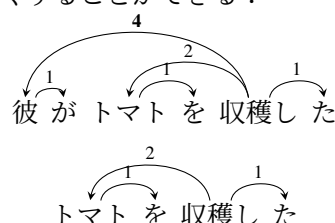
サプライザル 予測に関する処理負荷は、 n -gram

言語モデル ($n = 5$) での単語ごとの平均サプライザルで推定する。 n -gram は Kneser–Ney のバックオフスムージング [36] を用いた²⁾。

4 結果・考察

8つの言語と1,000のベースライン言語の記憶と予測の推定処理負荷の結果を図2に示す。ベースライン言語については、カーネル密度推定の結果をプロットした。 x 軸は平均依存関係長、 y 軸は平均サプライザルを表す。平均依存関係長と平均サプライザルは、比較のため、それぞれで z 値化した。

項省略の有無 日本語は動詞後置型の言語なので、省略可能な項を省略することにより、平均依存関係長を短くすることができる：



結果として、項省略がある場合、平均依存関係長は短くなった一方、項省略がない場合、すべての項が表出するため文脈情報が豊富になり、サプライザルは低下する傾向が見られた。

語順 Subj>Obj 語順に固定した場合は平均依存関係長が長くなる一方、次単語予測が容易になる傾向が見られた一方で、Obj>Subj 語順に固定すると、平均依存関係長は短縮されるものの、次単語予測が難しくなる傾向があった。これは、分析対象としたコーパスにおいて、目的語となる名詞句のほうが主語となる名詞句より長いことが多いため、主語を先に置くと依存関係が伸びやすい一方、目的語名詞句は主語名詞句より頻度が低く、文中でできる限り後ろに置かれる方が、文脈を活用して予測がしやすくなるためと考えられる。

自由語順である Id 語順は、この両者のバランスをとった地点に位置している。ここで、Swap 語順が記憶と予測の両観点において効率性が落ちる傾向にあることから、実際の言語は、単に語順に自由度があることが重要なのではなく、処理の効率性に形作

2) コーパス全体を文ごとすべてシャッフルした上、全体の2割を分離してテストデータとして用いる。残り8割分のデータで、交差検証 ($k = 5$) を行い、ハイパーパラメータである割引値 (discount) を $\{0.01, 0.02, \dots, 0.99\}$ の中からクロスエントロピーが平均で最小となる値として決定する。その上で、その8割分のデータで再度訓練した言語モデルを用い、テストデータにおける平均サプライザルを求める。

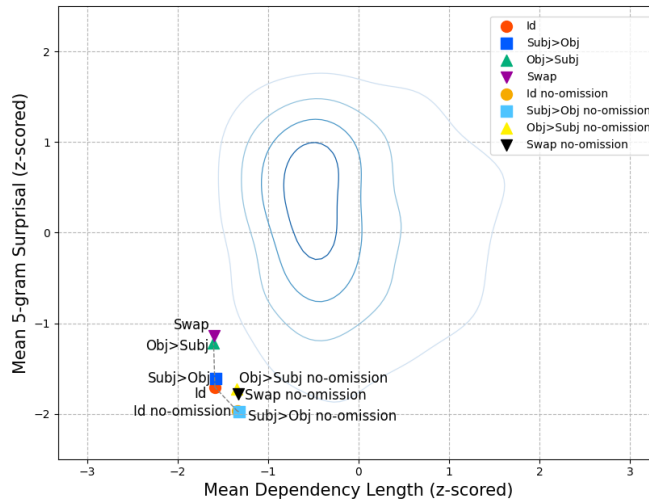


図2 記憶と予測における処理負荷の分布。青い分布は、ベースライン言語での処理負荷の推定値を表す。破線は、検証データ内でのパレート解 (Subj>Obj_{no-omission}, Id_{no-omission}, Id, Obj>Subj) の4言語を結んだパレートフロンティアを示す。

られた語順であることが示唆される。

パレート最適解 言語 L の処理負荷 $\Omega_L(\lambda)$ を、以下のように記憶負荷 (deplen) と予測負荷 (surprisal) の z 値の加重和として定義する：

$$\Omega_L(\lambda) := \lambda \cdot \text{deplen}_L + (1 - \lambda) \cdot \text{surprisal}_L \quad (0 \leq \lambda \leq 1)$$

ここで、 λ は、記憶と予測のどちらを重視するかを示す重みである。特定の λ に対して、 $\Omega_L(\lambda)$ が最小となる言語 L^* を「最も効率的な言語」とみなすとき、

$$L^* = \arg \min_{L \in \mathcal{L}} \Omega_L(\lambda)$$

各 λ の範囲で得られた結果は以下の通りである：

最も効率的な言語 L^*	λ の範囲
Subj>Obj _{no-omission}	[0.00, 0.11]
Id _{no-omission}	[0.12, 0.52]
Id	[0.53, 0.97]
Obj>Subj	[0.98, 1.00]

まず、 $\lambda \in [0.53, 0.98]$ の範囲で、**自由語順かつ項省略あり**の Id 言語が最適解となった。これは、自由語順と項省略の組み合わせが、特に記憶の処理負荷を小さくすることを優先した状態で、処理負荷を最小化するのに寄与することを示唆する。

一方、 λ が極端に大きい (記憶負荷を重視) 場合は、項省略があり語順が目的語・主語の順に固定された Obj>Subj 言語が最適である。ただし、世界の言語の大半は主語を文頭に置く強い傾向があることが報告されており [37]、これには人間のコミュニケーション上の選好が関与していると指摘されてい

る [38]。したがって、処理負荷のみだけではなく、情報伝達性や認知的要因も合わせて考慮するとこの結果は必ずしも維持されないかもしれない。

また、 $\lambda \in [0.12, 0.52]$ の範囲、すなわち、予測の処理負荷を重点的に考慮する値に設定した場合、項省略がない Id_{no-omission} 言語が、そしてさらに予測負荷を重視した場合には、語順を固定した Subj>Obj_{no-omission} 言語が最適解となった。興味深いことに、近年は次単語予測を最適化した大規模言語モデルが論文の文体にも影響を及ぼし始めている [39]。こうした「予測偏重型」モデルの普及により、将来的に日本語でも項省略が減り、Subj>Obj 語順への固定化が進むかもしれない。もちろん、実際の言語変化には社会的・文化的要因も複雑に関与するため、本研究の結果が直ちに言語変化を決定づけるわけではないものの、処理負荷という観点からは興味深い示唆を与えるだろう。

5 おわりに

本研究では、日本語における自由語順と項省略の併存が、記憶と予測の負荷のトレードオフを最適化している可能性を検証した。コーパスを用いた定量評価の結果、両現象が共存することで、特に記憶負荷の削減を優先した場合の最適解であることが示された。今後は、(i) 語順の自由度や項省略が日本語とは異なる性質を持つ他言語への適用を検討するとともに、(ii) 単純性だけではなく情報伝達性にも注目し、自由語順や項省略に加え、主題化や格標示といった文法現象を含むモデル化へ拡張する。

謝辞

本研究は JSPS 科研費 JP24H00087、JP21K00541、JST さきがけ JPMJPR21C2、国語研共同研究「計算言語学的手法による理論言語学の実証的な方法論の開拓」、国語研 E3P センターの支援を受けたものです。

参考文献

- [1] D. M. Eberhard, G. F. Simons, and C. D. Fennig. **Ethnologue: Languages of the World**. SIL International, Dallas, 27 edition, 2024.
- [2] T. F. Jaeger and H. Tily. On language ‘utility’: processing complexity and communicative efficiency. **WIREs Cognitive Science**, Vol. 2, No. 3, pp. 323–335, 2011.
- [3] C. Kemp, Y. Xu, and T. Regier. Semantic typology and efficient communication. **Annual Review of Linguistics**, Vol. 4, No. 1, pp. 109–128, 2018.
- [4] E. Gibson, R. Futrell, S. T. Piantadosi, I. Dautriche, K. Mahowald, L. Bergen, and R. Levy. How efficiency shapes human language. **Trends in Cognitive Sciences**, Vol. 23, No. 5, pp. 389–407, 2019.
- [5] N. Levshina. **Communicative Efficiency**. Cambridge University Press, 2022.
- [6] E. Bates and B. MacWhinney. Functionalist approaches to grammar. In **Language Acquisition: The State of the Art**, chapter 6, pp. 173–218. Cambridge University Press, 1982.
- [7] T. Givón. Markedness in grammar: Distributional, communicative and cognitive correlates of syntactic structure. **Studies in Language**, Vol. 15, No. 2, pp. 335–370, 1991.
- [8] J. A. Hawkins. **A Performance Theory of Order and Constituency**. Cambridge University Press, 1994.
- [9] M. Haspelmath. Parametric versus functional explanations of syntactic universals. **The Limits of Syntactic Variation**, Vol. 132, pp. 75–107, 2008.
- [10] S. T. Piantadosi, H. Tily, and E. Gibson. The communicative function of ambiguity in language. **Cognition**, Vol. 122, No. 3, pp. 280–291, 2012.
- [11] N. Zaslavsky, C. Kemp, T. Regier, and N. Tishby. Efficient compression in color naming and its evolution. **PNAS**, Vol. 115, No. 31, pp. 7937–7942, 2018.
- [12] S. Kirby, M. Tamariz, H. Cornish, and K. Smith. Compression and communication in the cultural evolution of linguistic structure. **Cognition**, Vol. 141, pp. 87–102, 2015.
- [13] M. Hahn, D. Jurafsky, and R. Futrell. Universals of word order reflect optimization of grammars for efficient communication. **PNAS**, Vol. 117, No. 5, pp. 2347–2353, 2020.
- [14] K. Kajikawa, Y. Kubota, and Y. Oseki. Is structure dependence shaped for efficient communication?: A case study on coordination. In **CoNLL 2024**, pp. 291–302, 2024.
- [15] K. Hale. Remarks on Japanese phrase structure: Comments on the papers on Japanese syntax. In **MIT Working Papers in Linguistics**, Vol. 2, pp. 185–203, 1980.
- [16] S. Oku. LF copy analysis of Japanese null arguments. In **CLS**, Vol. 34, pp. 299–314, 1998.
- [17] M. Saito. On the role of selection in the application of merge. In **NELS**, Vol. 33, pp. 323–345, 2002.
- [18] D. Takahashi. Noun phrase ellipsis. In **The Oxford Handbook of Japanese Linguistics**. Oxford University Press, 11 2008.
- [19] M. Asahara, H. Kanayama, T. Tanaka, Y. Miyao, S. Uematsu, S. Mori, Y. Matsumoto, M. Omura, and Y. Murawaki. Universal Dependencies version 2 for Japanese. In **LREC 2018**, 2018.
- [20] Y. Ishizuki, T. Kuribayashi, Y. Matsubayashi, R. Sasano, and K. Inui. To drop or not to drop? predicting argument ellipsis judgments: A case study in Japanese. In **LREC-COLING 2024**, pp. 16198–16210, 2024.
- [21] E. Gibson. The dependency locality theory: A distance-based theory of linguistic complexity. In **Image, Language, Brain: Papers from the First Mind Articulation Project Symposium**, pp. 95–126. The MIT Press, 2000.
- [22] S. Isono. Category locality theory: A unified account of locality effects in sentence comprehension. **Cognition**, Vol. 247, p. 105766, 2024.
- [23] J. Hale. A probabilistic Earley parser as a psycholinguistic model. In **NAACL 2001**, 2001.
- [24] R. Levy. Expectation-based syntactic comprehension. **Cognition**, Vol. 106, No. 3, pp. 1126–1177, 2008.
- [25] R. Futrell, E. Gibson, and R. P. Levy. Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. **Cognitive Science**, Vol. 44, No. 3, p. e12814, 2020.
- [26] H. Yamashita and F. Chang. “Long before short” preference in the production of a head-final language. **Cognition**, Vol. 81, No. 2, pp. B45–B55, 2001.
- [27] N. Orita. Predicting Japanese scrambling in the wild. In **CMCL 2017**, pp. 41–45, 2017.
- [28] N. J. Smith and R. Levy. The effect of word predictability on reading time is logarithmic. **Cognition**, Vol. 128, No. 3, pp. 302–319, 2013.
- [29] C. Shain, C. Meister, T. Pimentel, R. Cotterell, and R. Levy. Large-scale evidence for logarithmic effects of word predictability on reading time. **PNAS**, Vol. 121, No. 10, p. e2307876121, 2024.
- [30] 石月由紀子, 栗林樹生, 松林優一郎, 大関洋平. 情報量に基づく日本語項省略の分析. 言語処理学会第 28 回年次大会, pp. 442–447, 2022.
- [31] M. Hahn, J. Degen, and R. Futrell. Modeling word and morpheme order in natural language as an efficient tradeoff of memory and surprisal. **Psychological Review**, Vol. 128, pp. 726–756, December 2021.
- [32] D. Gildea and T. F. Jaeger. Human languages order information efficiently. **arXiv**, 2015.
- [33] M. Hahn and Y. Xu. Crosslinguistic word order variation reflects evolutionary pressures of dependency and information locality. **PNAS**, Vol. 119, No. 24, p. e2122604119, 2022.
- [34] K. Maekawa, M. Yamazaki, T. Ogiso, T. Maruyama, H. Ogura, W. Kashino, H. Koiso, M. Yamaguchi, M. Tanaka, and Y. Den. Balanced corpus of contemporary written Japanese. **Language resources and evaluation**, Vol. 48, pp. 345–371, 2014.
- [35] D. Gildea and D. Temperley. Optimizing grammars for minimum dependency length. In **ACL2007**, pp. 184–191, Prague, Czech Republic, 2007.
- [36] R. Kneser and H. Ney. Improved backing-off for M-gram language modeling. In **ICASSP**, pp. 181–184, 1995.
- [37] M. S. Dryer. Order of subject, object and verb (v2020.4). In M. S. Dryer and M. Haspelmath, editors, **The World Atlas of Language Structures Online**. Zenodo, 2013.
- [38] R. Futrell, T. Hickey, A. Lee, E. Lim, E. Luchkina, and E. Gibson. Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. **Cognition**, Vol. 136, pp. 215–221, 2015.
- [39] K. Matsui. Delving into PubMed records: Some terms in medical writing have drastically changed after the arrival of ChatGPT. **medRxiv**, 2024.