

# CCG による日本語脳波データのモデリング

磯野 真之介<sup>1,3,4\*</sup> 梶川 康平<sup>1,3\*</sup> 杉本侑嗣<sup>2\*</sup> 浅原正幸<sup>3</sup> 大関洋平<sup>1</sup><sup>1</sup> 東京大学 <sup>2</sup> 大阪大学<sup>3</sup> 国立国語研究所 <sup>4</sup> 日本学術振興会特別研究員 (DC2)

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

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

masayu-a@ninja.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流のデータを用いた先行研究で示唆されている。そこで本研究では、日本語を対象とした新たな脳波データセットを構築した上で、CCG による指標が脳波の予測に寄与するのか検証する。統語操作の適用数および作業記憶の負荷に関する指標が、それぞれ事象関連電位を予測することを示す。

## 1 はじめに

言語学では、自然言語の文法には階層的な統語構造があると考えられてきた [1]。この仮説の下では、人間のリアルタイムの文理解では、統語構造が逐次的に構築されていることになる [2, 3, 4]。そのような文理解の過程をモデルできる文法理論として、言語学的に妥当な記述力を持ち、かつ逐次的な構造構築が可能な Combinatory Categorical Grammar (CCG) [5, 6] があり、行動・脳活動データからその妥当性が示されてきた。具体的には、fMRI による脳血流データ (英語) [7]、視線計測データ (日本語・英語) [8]、自己ペース読文データ (英語・日本語) [9, 10] が予測できることが示されている。

しかし、CCG に基づく指標が脳波データを予測できるのかはまだ明らかになっていない。また、これらの分析においては、CCG をもとにしつつも、構造構築の複雑さを捉える異なる指標が用いられている。文法操作の適用数に関する指標 [7, 8] や記憶負荷に関する指標 [9] が提案されており、どの指標が

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 ベースの言語モデル (GPT-2) のサプライザルも回帰モデルに含めた。結果、これらの指標を含めても、CCG に基づく指標が脳波データの予測に独立に寄与することが示された。具体的には、記憶負荷に関する指標が、統語ないし意味的情報統合に関わるとされる事象関連電位を予測することが示された。また、文法操作の適用数に関する複数の指標も、それぞれ異なる事象関連電位での説明力を示すが、同時に、これらの指標は、先行研究で主張されてきたような処理負荷として必ずしも解釈できるわけではないことも指摘する。

## 2 CCG による逐次的な文処理

人間の文理解は統語処理においても意味処理においても逐次的であるため (e.g., [2, 11])、逐次的な計算過程を厳密に記述できる文法理論が重要である。本研究では、記述的な妥当性と、逐次的な統語計算を実現できる CCG を用いて、CCG に基づくどの指標が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を検証する。具体的には、先行研究で提案されている 3 種類の指標 (各種の文法操作の適用数 [8]、節点数 [7]、記憶負荷 [9]) を検証する。

### 2.1 CCG

CCG では、各語彙項目にカテゴリが割り当てられている。カテゴリは  $S, NP$  などの基本カテゴリか、 $X/Y$  ないし  $X \backslash Y$  の形式 ( $X, Y$  はカテゴリ) の関数カ

\* 共同第一著者

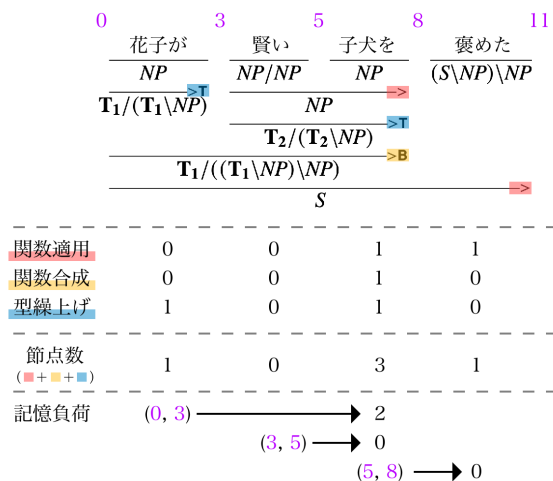


図1 CCGによる逐次的な構造構築過程の導出木と、各文節における処理の複雑さに関する指標の算出例。

テゴリである。これらのカテゴリに各種の操作を再帰的に適用することで、統語構造が作られる。主要な文法操作は以下の3種類である：

$$\begin{aligned}
 &\text{関数適用} && \text{型繰り上げ} \\
 &\left\{ \begin{array}{l} X/Y \quad Y \Rightarrow_{>} X \\ Y \quad X/Y \Rightarrow_{<} X \end{array} \right. && \left\{ \begin{array}{l} X \Rightarrow_{>T} T/(T/X) \\ X \Rightarrow_{<T} T/(T/X) \end{array} \right. \\
 &\text{関数合成} \\
 &\left\{ \begin{array}{l} X/Y \quad Y/Z \Rightarrow_{>B} X/Z \\ Y/Z \quad X/Y \Rightarrow_{<B} X/Z \end{array} \right.
 \end{aligned}$$

ここで  $X, Y$  は任意のカテゴリである。 $T$  は変数的カテゴリである。

## 2.2 予測に用いる指標

CCGでは、関数合成や型繰り上げを活用することで、構造構築がある程度逐次的にできる。この逐次的な統語構造から、処理の複雑さに関する次のような指標を導出できる（図1）。

**節点数** 各単語における操作の適用数（節点数）をその単語における処理コストと考えることができ、脳血流データを予測できることが示されている[7]。

**各種の文法操作の適用数** 節点数は各種の文法操作を区別せずに扱ってしまうが、それぞれの文法操作は読み時間に対する効果が異なることが示されている[8]。

**作業記憶の負荷** 心理言語学では、人間の作業記憶の限界に由来する処理負荷が存在することが長く指摘されている[12]。Category Locality Theory [9]では、CCG上の構成素が作業記憶から取り出される際に、それが構築されてからの時間経過に応じた処理

負荷がかかると提案されており、英語の自己ペース読文法による読み時間を予測できることが示されている。たとえば、図1では、「子犬を」の処理にあたって「花子が」を想起する必要があり、両者の距離（文字数ベースで2）が負荷となる。

## 3 実験

### 3.1 脳波実験手続き

**実験参加者** 本研究の脳波実験には42名の日本語母語話者（全員右利きの健常成人、女性22名、平均年齢20.5歳（SD=2.89））が参加した。参加者は全員正常な視力（矯正を含む）であった。

**刺激文と課題** 本研究で使用する刺激文は、BCCWJ-EyeTrack [13]と同様に、「現代日本語書き言葉均衡コーパス (Balanced Corpus of Contemporary Written Japanese: BCCWJ, [14])」にある様々なジャンルの中から新聞記事20件を使用した。各記事はランダムに提示され、各記事は文節ごとに PsychoPy [15, 16] を使用し、Rapid Serial Visual Presentationによって各文節は500ms間画面に提示された（各文節間には500ms何も表示されない時間帯を設けた）。また各記事の最後に質問が提示された。本実験では、参加者が各文節を読んでいる際の脳波活動のデータ収集を行った。脳波測定には、Brain Products製の64-ch装置 (BrainAmp)、64-ch脳波キャップを用い、脳波記録には、同社製の Brain Vision Recorder を使用した。

**前処理** オンラインでは、1000Hzでデータ収集を行い、オフラインで200Hzまでダウンサンプリングし、0.1-40Hz間でバンドパスフィルタを適用した。また独立成分分析 (ICA) によって瞬きなどの文処理に関連しないノイズの除去を行い（センサーごとの大きなノイズなども手動で除去した）、Rereferenceとして、デフォルト設定のaverageを使用した。-0.1-1000ms間でエポック化し、ベースライン修正として、-0.1-0.0ms間の脳波活動を使用した。

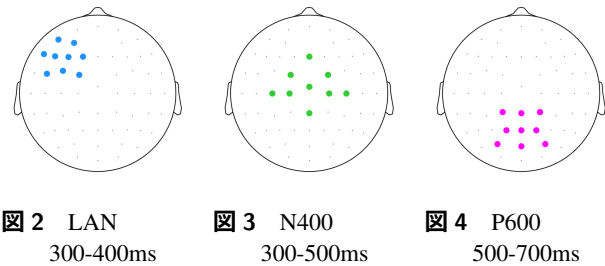
### 3.2 評価方法

以下の脳波解析には、MNE-Python (v1.9.0) [17]、Eelbrain (v0.40.1) [18]、R (v4.3.1) [19]を使用した。統計解析には前処理を施した41名分の脳波データを使用した。

**Spatio-temporal clustering tests** まず全センサーで計測された脳波に対して、spatio-temporal

clustering tests を行った。この検定では、0ms から 1000ms の区間の全センサーに対して CCG に基づく各説明変数がどのセンサーの、どのタイミングの脳波活動を説明するかを検証する。

**関心領域解析** 関心領域 (Region of interest, ROI) は、文処理に関する先行研究 [20, 21] より、LAN、N400、P600 を選択した。<sup>1)</sup> 時間幅は [20] を参考にした。また文頭・文末のトークンは取り除いた。



分析は線形混合効果モデルで行った。関心領域の脳波データを従属変数とし、ベースラインとなる説明変数に CCG に基づく説明変数のいずれかを加え、これらの固定効果と参加者ごとのランダム効果、および参加者・記事ごとのランダム切片をモデルに含めた [22]。lme4 の式ではモデルは以下のとおりである ( $X$  が CCG に基づく説明変数のいずれか。ベースライン説明変数の詳細は付録の表 3 に示す。また分析に使用した説明変数間の相関も付録の図 6 のとおりである。):

$$\begin{aligned} ROI \sim & \text{word\_length} + \text{word\_freq} + \text{sentid} + \text{bunsetsupos} \\ & + \text{numwords} + \text{topdown} + \text{bottomup} \\ & + \text{gpt2\_med} + \text{ndep} + \text{deplen} + X + \\ & + (1 + \text{word\_length} + \text{word\_freq} + \text{sentid} \\ & + \text{bunsetsupos} + \text{numwords} + \text{topdown} + \text{bottomup} \\ & + \text{gpt2\_med} + \text{ndep} + \text{deplen} + X || \text{subject}) \\ & + (1 | \text{section\_num}) \end{aligned}$$

最初のモデルを当てはめたあと、標準偏差の絶対値が 3 以上のデータポイントを取り除き、再度モデルを当てはめた [23]。

## 4 結果と考察

### 4.1 Spatio-temporal clustering tests

Spatio-temporal clustering tests の主な結果は、表 1 と図 5 のとおりである。関数適用、型繰り上げ、節

1) 先行研究とはモニターが異なるため、センサー位置は近似した領域を用いた。

点数、記憶負担では時空間クラスターが確認された。関数適用と節点数は、頭頂部付近のセンサーで

表 1 Spatio-temporal clustering tests の主要な結果

	time window	$p_{cluster}$
関数適用	150 - 635ms	<b><math>p=0.0017</math></b>
関数合成	not significant	
型繰り上げ	140 - 575ms	<b><math>p=0.0009</math></b>
節点数	10 - 730ms	<b><math>p&lt;0.0001</math></b>
記憶負担	484 - 715ms	<b><math>p=0.0301</math></b>
	725 - 915ms	<b><math>p=0.0414</math></b>

P600 も含む時空間上のデータの説明を示唆しており、CCG の各説明変数が文処理レベルで効果があることが示唆される。また、型繰り上げは、左前頭部から左中側頭部での脳波活動を予測しており、LAN 効果を説明していると思われる。記憶負担に関しては、中心部の遅いタイミングでの効果を示していることが確認された。なお、合成関数は、統計的に有意な時空間クラスターが確認されなかった。

### 4.2 関心領域解析

各線形混合効果モデルにおける CCG に基づく説明変数の係数 ( $\beta$ ) と  $p$  値を表 2 に示す。型繰り上げが LAN で有意な正の効果を、節点数が N400 で有意な正の効果を、記憶負担が P600 で有意な正の効果を示した。また関数適用が P600 で境界有意な正の効果を示した。一方、関数合成はいずれの関心領域でも有意な効果を示さなかった。これらの結果は spatio-temporal clustering tests の結果とおおむね整合する。

### 4.3 考察

**LAN** LAN は、格一致や文法的一致などの形態統語的な違反や、統語的予測への違反に対応して観察されている [24, 25]。今回の分析では型繰り上げが LAN の関心領域で正の効果（つまり LAN が弱い）ことが示された。本研究で採用している CCG の統語構造では、型繰り上げは名詞句にのみかかる操作であることから、LAN が主に観察されるのは用言であることを示しているとも解釈できる。この解釈は、名詞句の格により統語構造が予測され、用言により“答え合わせ”がなされる日本語の文理解のあり方とも整合する。

**N400** N400 は語の予測しにくさに対応して観察されることが知られている [20]。今回の分析では節



表2 各脳波成分を従属変数とした線形混合効果モデリングの結果。太字は  $p < 0.05$ 、斜体は  $p < 0.1$ 。

	LAN		N400		P600	
	$\beta$	$p$	$\beta$	$p$	$\beta$	$p$
関数適用	$-2.0 \times 10^{-09}$	0.9509	$2.1 \times 10^{-08}$	0.4760	$8.2 \times 10^{-08}$	0.0566
関数合成	$1.1 \times 10^{-08}$	0.7535	$3.9 \times 10^{-09}$	0.9088	$-5.9 \times 10^{-08}$	0.1883
型繰り上げ	$9.8 \times 10^{-08}$	<b>0.0002</b>	$-4.2 \times 10^{-09}$	0.8541	$-1.2 \times 10^{-07}$	<b>0.0007</b>
節点数	$8.7 \times 10^{-08}$	0.1624	$1.3 \times 10^{-07}$	<b>0.0132</b>	$8.5 \times 10^{-08}$	0.2641
記憶負荷	$-7.5 \times 10^{-09}$	0.7549	$3.1 \times 10^{-08}$	0.1362	$6.6 \times 10^{-08}$	<b>0.0403</b>

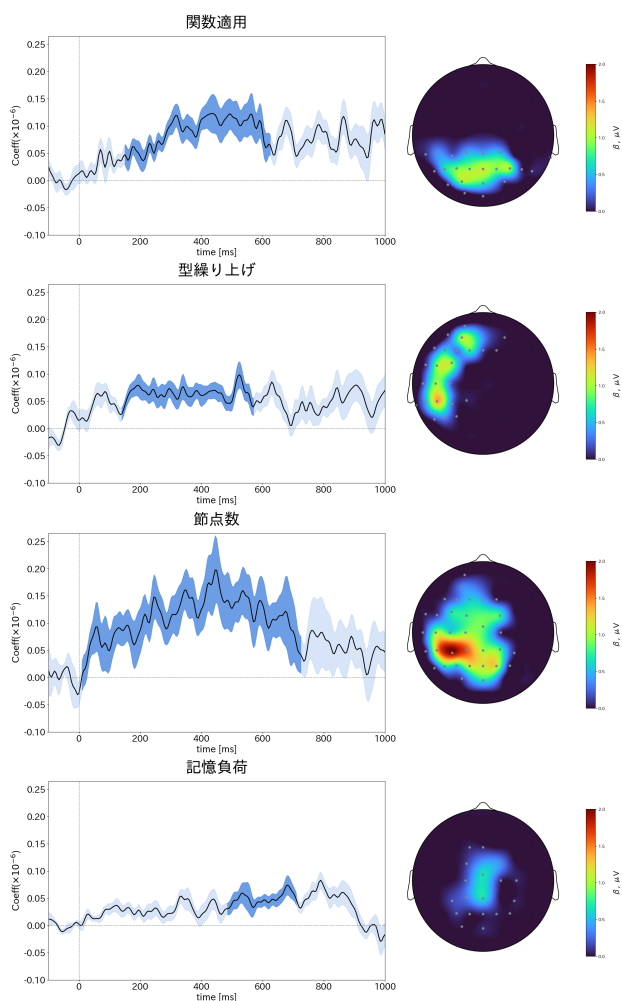


図5 Spatio-temporal clustering tests の結果 (上から順に、関数適用、型繰り上げ、節点数、記憶負荷)。プロットされた値は、回帰係数の推定値 (黒色) と 95%信用区間 (水色) を示す。濃い水色の部分が検定に基づく統計的有意性を示す。

点数が N400 の関心領域で正の効果 (つまり N400 が弱い) を示した。これは、節点数が多いトークンほど予測可能性が高いことを示唆している。節点数が多いということは、そのトークンと統合される情報が多いということであり、その分トークンの予測可能性が高いと考えることができる [10]。

**P600** P600 がどのような処理を反映しているのかは議論があるが、一致の違反、ガーデンパス文、

フィラー・ギャップ依存関係などに対応して観察されており [26, 27, 28]、統語的情報と意味的情報の統合の困難さを反映しているともいわれる [29]。今回の分析では、記憶負荷が P600 の関心領域で有意な正の効果を示したほか、関数適用が境界有意な正の効果を示し、これらは統語的な統合処理の困難さに対応しているとも考えられる。一方、型繰り上げは P600 の関心領域で有意な負の効果を示した。これは LAN と同様に、P600 が用言において生じやすいことを示唆する。

## 5 おわりに

本研究では、日本語を対象にした新たな脳波データセットを構築し、CCG における操作の回数と CCG に基づく記憶負荷の指標を使用して解析・検証した。扱った CCG の操作のうち、関数適用、関数合成、節点数がそれぞれ統語処理に関連した事象関連電位での説明力を示した。また記憶負担も統語・意味的情報統合に関わる事象関連電位における効果を示し、CCG に基づく指標が脳波データの予測に寄与することを示した。

## 謝辞

脳波データを収集していただいた小林由紀氏、中村梓甫氏、原田宥都氏、深津聡世氏、森田早織氏、山下陽一郎氏、また実験に参加していただいた東京大学の学生の方々に感謝致します。本研究は、JST さきがけ JPMJPR21C2 および JSPS 科研費 24H00087 の支援を受けたものです。

## 参考文献

- [1] Noam Chomsky. **Syntactic Structures**. Mouton, 1957.
- [2] Michael K. Tanenhaus, Michael J. Spivey-Knowlton, Kathleen M. Eberhard, and Julie C. Sedivy. Integration of visual and linguistic information in spoken language comprehension. **Science**, Vol. 268, No. 5217, pp. 1632–1634, 1995.
- [3] Jonathan Brennan, Yuval Nir, Uri Hasson, Rafael Malach, David J. Heeger, and Liina Pykkänen. Syntactic structure

- building in the anterior temporal lobe during natural story listening. **Brain and Language**, Vol. 120, No. 2, pp. 163–173, 2012. The Neurobiology of Syntax.
- [4] Jonathan R. Brennan, Edward P. Stabler, Sarah E. Van Wagenen, Wen-Ming Luh, and John T. Hale. Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. **Brain and Language**, Vol. 157–158, pp. 81–94, 2016.
  - [5] Mark Steedman. **The syntactic process**. MIT Press, 2000.
  - [6] 戸次大介. 日本語文法の形式理論. くろしお出版, 東京, 2010.
  - [7] Miloš Stanojević, Jonathan R. Brennan, Donald Dunagan, Mark Steedman, and John T. Hale. Modeling structure-building in the brain with CCG parsing and large language models. **Cognitive Science**, Vol. 47, No. 7, p. e13312, 2023.
  - [8] Kohei Kajikawa, Ryo Yoshida, and Yohei Oseki. Dissociating syntactic operations via composition count. In **CogSci 2024**, 2024.
  - [9] Shinnosuke Isono. Category Locality Theory: A unified account of locality effects in sentence comprehension. **Cognition**, Vol. 247, p. 105766, 2024.
  - [10] 磯野真之介, 梶川康平, 大関洋平. 日本語大規模読み時間コーパスにおける記憶の負荷の CCG によるモデリング. 日本言語学会第 169 回大会, 2024.
  - [11] Yuki Kamide, Christoph Scheepers, and Gerry T. M. Altmann. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. **Journal of Memory and Language**, Vol. 49, No. 1, pp. 133–156, 2003.
  - [12] Edward Gibson. Linguistic complexity: locality of syntactic dependencies. **Cognition**, Vol. 68, No. 1, pp. 1–76, 1998.
  - [13] Masayuki Asahara, Hajime Ono, and Edson T. Miyamoto. BCCWJ-EyeTrack. **Gengo Kenkyu**, Vol. 156, pp. 67–96, 2019.
  - [14] Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den. Balanced corpus of contemporary written Japanese. **Language Resources and Evaluation**, Vol. 48, No. 2, pp. 345–371, 2014.
  - [15] Jonathan W. Peirce. Psychopy–psychophysics software in python. **Journal of neuroscience methods**, Vol. 162, No. 1–2, pp. 8–13, 2007.
  - [16] Jonathan W. Peirce. Generating stimuli for neuroscience using psychopy. **Frontiers in Neuroinformatics**, Vol. 2:10, , 2009.
  - [17] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti Hämäläinen. Meg and eeg data analysis with mne-python. **Frontiers in Neuroscience**, Vol. 7, , 2013.
  - [18] Christian Brodbeck, Proloy Das, Marlies Gillis, Joshua P Kulasingham, Shohini Bhattachali, Phoebe Gaston, Philip Resnik, and Jonathan Z Simon. Eelbrain, a python toolkit for time-continuous analysis with temporal response functions. **eLife**, Vol. 12, p. e85012, nov 2023.
  - [19] RStudio Team. **RStudio: Integrated Development Environment for R**. RStudio, PBC., Boston, MA, 2020.
  - [20] Stefan L. Frank, Leun J. Otten, Giulia Galli, and Gabriella Vigliocco. The erp response to the amount of information conveyed by words in sentences. **Brain and Language**, Vol. 140, pp. 1–11, 2015.
  - [21] John Hale, Chris Dyer, Adhiguna Kuncoro, and Jonathan Brennan. Finding syntax in human encephalography with beam search. In Iryna Gurevych and Yusuke Miyao, editors, **Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 2727–2736, Melbourne, Australia, July 2018. Association for Computational Linguistics.
  - [22] Dale J. Barr, Roger Levy, Christoph Scheepers, and Harry J. Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. **Journal of Memory and Language**, Vol. 68, No. 3, pp. 255–278, 2013.
  - [23] R. H. Baayen, Douglas J. Davidson, and Douglas M. Bates. Mixed-effects modeling with crossed random effects for subjects and items. **Journal of Memory and Language**, Vol. 59, No. 4, pp. 390–412, 2008.
  - [24] Thomas F. Münte, Hans-Jochen Heinze, and George R Mangun. Dissociation of brain activity related to syntactic and semantic aspects of language. **Journal of Cognitive Neuroscience**, Vol. 5, No. 3, pp. 335–344, 07 1993.
  - [25] Masataka Yano. Predictive processing of syntactic information: evidence from event-related brain potentials. **Language, Cognition and Neuroscience**, Vol. 33, No. 8, pp. 1017–1031, 2018.
  - [26] Lee Osterhout and Linda A. Mobley. Event-related brain potentials elicited by failure to agree. **Journal of Memory and Language**, Vol. 34, No. 6, pp. 739–773, 1995.
  - [27] Edith Kaan and Tamara Y Swaab. Electrophysiological evidence for serial sentence processing: a comparison between non-preferred and ungrammatical continuations. **Cognitive Brain Research**, Vol. 17, No. 3, pp. 621–635, 2003.
  - [28] Colin Phillips, Nina Kazanina, and Shani H. Abada. Erp effects of the processing of syntactic long-distance dependencies. **Cognitive Brain Research**, Vol. 22, No. 3, pp. 407–428, 2005.
  - [29] Angela D. Friederici. The brain basis of language processing: From structure to function. **Physiological Reviews**, Vol. 91, pp. 1357–1392, 2011.
  - [30] Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. Stanza: A python natural language processing toolkit for many human languages. In Asli Celikyilmaz and Tsung-Hsien Wen, editors, **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations**, pp. 101–108, Online, July 2020. Association for Computational Linguistics.
  - [31] Kei Sawada, Tianyu Zhao, Makoto Shing, Kentaro Mitsui, Akio Kaga, Yukiya Hono, Toshiaki Wakatsuki, and Koh Mitsuda. Release of pre-trained models for the Japanese language. In **Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)**, pp. 13898–13905, 2024.

## A 説明変数

表3 本研究で用いたベースライン説明変数

変数名	型	概要
word_length	int	文節の文字数
word_freq	num	文節の単語頻度の対数平均
sentid	int	記事内の文の位置
bunsetsupos	int	文内の文節の位置
numwords	int	各句に構成される文節数
topdown	int	Stanza [30] による constituency parse をトップダウンに走査したときの節点数
bottomup	int	Stanza [30] による constituency parse をボトムアップに走査したときの節点数
gpt2_med	num	huggingface で実装されている rinna による日本語 GPT-2 (medium) [31] <sup>2)</sup> で算出したサプライズ値
ndep	int	先行文脈から当該文節にかかる係り受け関係の数
deplen	int	先行文脈から当該文節にかかる係り受け関係の距離の和
subject	factor	参加者番号
section_num	factor	記事番号

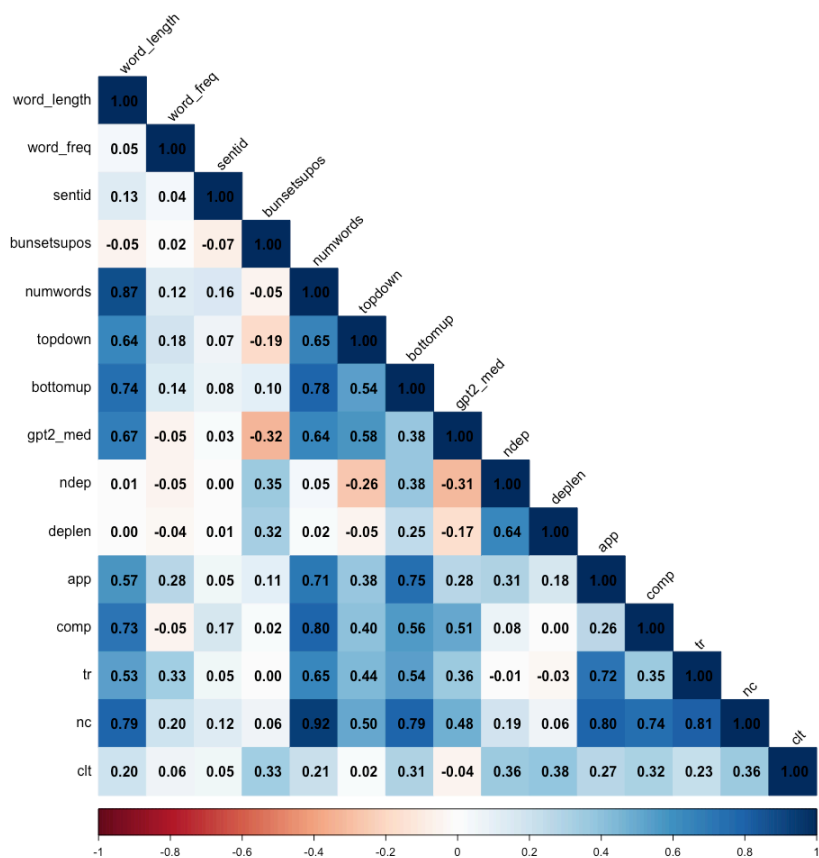


図6 説明変数間の相関。app は関数適用、comp は関数合成、tr は型繰り上げ、nc は接点数、clt は作業記憶の負荷。それ以外の説明変数については表3を参照。