

# 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

磯野真之介<sup>1,2</sup> 梶川康平<sup>1</sup> 大関洋平<sup>1</sup>

<sup>1</sup> 東京大学

<sup>2</sup> 日本学術振興会特別研究員 (DC2)

## 概要

逐次的な文理解においては、統語的關係にある語句の線形位置が離れている場合、作業記憶への負荷（局所性効果）が生じると考えられる。予測される局所性効果は、採用する文法理論によって異なるが、Isono (2024) は、組合せ範疇文法 (CCG) を採用することで、動詞先置型言語と後置型言語における局所性効果のパターンを統一的に説明しうることを示し、さらに英語コーパスの読み時間を予測できることを統計的に示した。これを踏まえ本研究では、日本語のコーパスに自己ペース読文による読み時間を付した大規模データセット (BCCWJ-SPR2) を用いて、CCG が日本語における局所性効果を予測できるかを検証した。線形混合効果モデルによる分析の結果、CCG に基づく距離の効果がスピルオーバー領域において観察され、この効果は従来の主流である依存文法に基づく距離の効果よりも安定していることが示された。

## 1 背景

文が持つ階層的な構造は、外在化された形式では直接観察することができず、聞き手ないし読み手は線形に並んだ語句からそれを復元しなければならない。その際、統語的に直接関係のある語句が、線形順序では離れた位置にあることがある。たとえば(1a)は英語の目的語関係節であるが、従属節の動詞 *attacked* の目的語に相当する *who* は、*attacked* に隣接しておらず、間に主語の *the senator* が挟まっている。

(1) (cf. Gibson, 2000)

- a. The reporter who the senator *attacked* disliked the editor.
- b. The reporter who *attacked* the senator disliked the editor.

このような文を逐次的に（左から右に）理解するときには、先行する語句を一旦作業記憶に保持して後に想起する必要があり、その際の負荷は、想起の対象（この例では *who*）が想起すべき地点（*attacked*）から離れているほど高いと考えられる (Gibson, 2000; Lewis & Vasishth, 2005)。こうした線形距離に基づく処理負荷は**局所性効果 (locality effect)** と呼ばれる (Lewis et al., 2006)。英語の目的語関係節は、対応する主語関係節(1b) に比べ読み時間が長いことが知られ、局所性効果はそれに対する有力な説明である (Bartek et al., 2011; Grodner & Gibson, 2005; King & Just, 1991; Staub, 2010; cf. Levy, 2008)。

局所性効果を説明する文理解モデルは様々なものが提唱されてきたが (Demberg et al., 2013; Futrell et al., 2020; Gibson, 2000; Lewis & Vasishth, 2005)、少なくとも2つの重要な経験的問題があった (Isono, 2024)。第一に、動詞後置型言語では、項と動詞の間で局所性効果がほとんど見られない (Isono & Hirose, 2023; Konieczny, 2000; Levy & Keller, 2013; Nakatani & Gibson, 2010; Vasishth & Lewis, 2006; cf. Safavi et al., 2016)。これは、実験で項と動詞の距離を伸ばすために他の項や付加詞を追加すると、動詞が予測しやすくなり、局所性効果を相殺するからだと考えられる (Konieczny & Döring, 2003; Levy, 2008; Vasishth & Lewis, 2006)。しかし、予測の効果が局所性効果を上回る原理的な理由は与えられてこなかったし、動詞の予測におよそ差がないと考えられる構文でも、やはり局

所性効果は見られない (Isono & Hirose, 2023; Nakatani & Gibson, 2010)。第二に、英語に限っても、幅広い構文に当てはまる (broad-coverage) 局所性効果の理論は得られていない。局所性効果の代表的理論である **dependency locality theory (DLT)** (Gibson, 2000) は複数の研究で英語コーパスの読み時間との関係が検証されてきたが、DLT の一般的な妥当性を示す結果は得られていない (Demberg & Keller, 2008; Dotlačil, 2021; Isono, 2024)。\*1

これに対し Isono (2024) は、理論の予測と観察の不一致の原因を、仮定すべき文法理論に求めた。従来のモデルでは、DLT が依存文法、もう一つの代表的な理論である Lewis and Vasishth (2005) のモデルが X バー理論を仮定していた。Isono は新たに、組合せ範疇文法 (CCG) (Steedman, 2000) に基づいて作業記憶への負荷を予測するモデル — **category locality theory (CLT)** — を提唱し、英語コーパス (Futrell et al., 2021) の読み時間を予測することおよび、動詞後置型言語における局所性効果の欠如を予測できることを示した。ただし、後者は概念的な提示にとどまっており、コーパスのデータで検証されていない。また、CLT が動詞後置型言語で負荷の存在を積極的に予測する場面についても議論されていない。これを踏まえ、本研究は、大規模な日本語の読み時間データセットである BCCWJ-SPR2 (Asahara, 2022) で CLT の予測を検証し、肯定的な結果を得た。以下、2 節で CLT を紹介したうえで、3 節でコーパス分析の結果を報告し、4 節で議論する。

## 2 CCG に基づく局所性効果の理論

CCG (Steedman, 2000) の特徴の一つは文理解との親和性である。CCG では構成素の概念が柔軟であり、逐次的な文理解で解析器が遭遇する未完成の句の多くに対しても構成素の地位と論理形式を与えられる。

CCG では各語彙および構成素が範疇を持ち、その範疇に従って規則を再帰的に適用することで派生が進む。主要な規則を(2)に示す。範疇は、 $S, NP, N$  といった基本的範疇と、 $X, Y$  を範疇として  $X/Y$  もしくは  $X \setminus Y$  の形式を持つ関数的範疇からなる。関数的範疇はその名の通り項をとって新たな範疇を返すもので、その基本的振る舞いは**関数適用規則** ( $>$ ,  $<$ ) により規定される。CCG には関数適用規則に加えていくつかの規則群があるが、核となるのは**関数合成規則** ( $>B$ ,  $<B$  等) と**型繰り上げ規則** ( $>T$ ,  $<T$ ) である。\*2

$$(2) \quad \frac{X/Y \quad Y}{X} > \quad \frac{Y \quad X \setminus Y}{X} < \quad \frac{X/Y \quad Y/Z}{X/Z} >^B \quad \frac{Y \setminus Z \quad X \setminus Y}{X \setminus Z} <^B \quad \frac{X}{Y/(Y \setminus X)} >^T \quad \frac{X}{Y \setminus (Y/X)} <^T$$

関数適用規則のみでは、文脈自由文法相当の表現力があり、たとえば SVO 構文に対して(3a)のような派生を与えられるが、関数合成・型繰り上げ規則の使用により、同じ文に対して左枝分かれの構造(3b)も与えられる。本発表では詳述しないが、CCG の構成素には論理形式が付随するため、解析器が *Pat likes* のような部分的入力を解釈できることが、CCG を仮定することで難なく説明できる。*Pat likes* のような非伝統的構成素は、右方繰り上げや目的語関係節などの派生にも必要であり、文理解とは独立に動機づけられている。また、左枝分かれの派生であっても、それに伴って生成される論理形式は右枝分かれの階層構造を持っており、束縛条件などはそちらで説明される。

$$(3) \quad \begin{array}{c} \text{a.} \quad \text{Pat} \quad \text{likes} \quad \text{Kim} \\ \hline \text{NP} \quad (S \setminus \text{NP}) / \text{NP} \quad \text{NP} \\ \hline \text{S} \setminus \text{NP} \quad > \\ \hline \text{S} \quad < \end{array} \quad \begin{array}{c} \text{b.} \quad \text{Pat} \quad \text{likes} \quad \text{Kim} \\ \hline \text{NP} \quad (S \setminus \text{NP}) / \text{NP} \quad \text{NP} \\ \hline \text{S} / (S \setminus \text{NP}) \quad >^T \\ \hline \text{S} / \text{NP} \quad >^B \\ \hline \text{S} \quad > \end{array}$$

CLT (Isono, 2024) は、CCG 上の構成素が記憶の単位であり、ある構成素の形成から、次にそれが規則の適用対

\*1 Shain et al. (2016) は DLT が英語コーパスの読み時間を予測できると報告したが、後に結果を見直している (Shain & Schuler, 2018)。

\*2 型繰り上げについては自由に認めると島の制約が破られることから近年では語彙的な現象とされる (Steedman, 2023)。しかし型繰り上げの語彙化は過剰生成も引き起こすため、本研究は型繰り上げは統語規則でもあるとの立場を取る (語彙的に型繰り上げされている項目があることは排除しない)。これは本研究で用いる *depccg* (Yoshikawa et al., 2017) の実装とも合致する。詳細は Isono (2024) を参照。



ブ上で被験者に読ませたもので、PB レジスターは書籍 83 サンプル、10,075 文、84,736 文節からなる。サンプルごとの内容理解課題に不正解だった試行と、80 ミリ秒以下および 3000 ミリ秒以上の読み時間はあらかじめ取り除かれている。本研究ではデータポイントが 1000 個以下の被験者も除外し、336 人分を対象としている。

### 3.2 CCG に基づく予測子の算出

分析対象の文は depccg<sup>\*4</sup> (Yoshikawa et al., 2017) で解析したうえで、独自に最も逐次的な木に変換した。CLT のコストは、この変換後の木をもとに、各構成素が形成されてから次に規則適用の対象となるまでの距離によって算出される。本検証では距離の算出方法として、(i) 文節ごとに 1、(ii) 文字ごとに 1、の 2 種類を考慮した。ただし文字ごとの場合も、コスト自体は文節に対して割り当てた。たとえば (5) では、「を」の入力時に「記者が」と「有名な議員を」が合成される。その際「記者が」が形成されてから「を」を含む文節である「議員を」に到達するまでには「有名な」が処理されている。「有名な」は 1 文節 3 文字であるため、「を」のコストは文節基準で 1、文字基準で 3 となる。これが「を」を含む文節である「議員を」のコストに算入される。「議員を」では他にコストのかかる規則適用はないので、結局「議員を」の CLT コストも同様である。2 種類の基準を考慮したのは、CLT コストにおける“距離”は理論上、間に挟まる要素の処理負荷によって定まるべきだが、それを厳密に決定する自明な方法がないからである。本データセットでは刺激呈示が文節ごとであったことと、文字数は一般に読み時間の単純かつ強力な予測子であることを考慮して、文節数と文字数に基づいて CLT コストを算出した。

CCG に基づくもう一つの予測子として、規則適用の回数も考慮に入れた。CCG における規則適用の回数も、読み時間や脳活動を予測できることが報告されている (梶川ほか, 2023; Kajikawa et al., to appear; Stanojević et al., 2023)。本分析では、文節をまたぐ二枝分かれの規則適用の回数を予測子に含めることとした。この値は CLT コストと関連しやすく、モデルに含めることで CLT が捉えようとする距離の効果を規則適用の回数の効果から分離できる。

### 3.3 統計モデリング

分析には線形混合効果モデル (Baayen et al., 2008) を用いた。従属変数は、対数変換した文節ごとの読み時間 (文頭・文末を除く) とした。説明変数としては以下を含めた。

- 文節の低級な属性に基づくもの：サンプル内の文の位置・文内の文節の位置・文字数・短単位数
- サプライザル (Hale, 2001; Levy, 2008) に基づくもの：短単位の unigram サプライザルの和 (国語研日本語ウェブコーパス (Asahara et al., 2014) の n-gram 頻度表<sup>\*5</sup>をもとに計算)・GPT-2<sup>\*6</sup> (Radford et al., 2019) によるサブワードごとのサプライザルの和
- 依存文法に基づくもの：その文節で解決される依存関係の距離 (これも文節基準、文字基準の 2 種類を算出した) (Gibson, 2000) とその数 (浅原ほか, 2019; Konieczny & Döring, 2003)
- CCG に基づくもの：CLT コスト、規則適用の回数 (上述)

すべての予測子は  $z$  値化し、固定効果としてモデルに含めた。また、依存文法および CCG に基づく予測子は被験者ごとのランダム効果としても含め (cf. Barr et al., 2013)、被験者およびサンプルごとのランダム切片も含めた。GPT-2 のサプライザルは、自動で行われるトークナイゼーション (サブワード化) が文節をまたぐ場合があり、その場合はそれらの文節を分析から除外した。データが巨大 (約 480 万データポイント) であるため、20~21 サンプルごとに 4 分割し、それぞれにモデルを適用した。モデルが収束するまでランダム効果を分散の小さいものから順に取り除いた (Bates et al., 2015)。その後、モデルからの標準残差の絶対値が 2.5 を超えるデータポイントを外れ値として取り除き、再度モデルを適用した (Baayen et al., 2008)。

<sup>\*4</sup> <https://github.com/masashi-y/depccg>

<sup>\*5</sup> <https://www.gsk.or.jp/catalog/gsk2020-c/>

<sup>\*6</sup> <https://huggingface.co/rinna/japanese-gpt2-medium>

	文節基準					文字基準				
	係数	1-20	21-41	42-62	63-83	係数	1-20	21-41	42-62	63-83
依存関係の距離	-0.002	-				0.001				+
スピルオーバー	0.001			+	+	0.004	+		+	+
依存関係の数	-0.007	-	-	-	-	-0.008	-	-	-	-
スピルオーバー	-0.010	-	-	-	-	-0.012	-	-	-	-
CLT コスト	-0.001					0.000			+	
スピルオーバー	0.002	+	+	+	+	0.003	+	+	+	+
CCG 規則適用の回数	-0.002	-	-	-	-	-0.002	-	-	-	-
スピルオーバー	-0.009	-	-	-	-	-0.009	-	-	-	-

表1 線形混合効果モデルによる分析の結果の概要。「係数」はサンプル 1-20 を対象としたモデルの係数。横軸の数字はサンプル。+, - はそれぞれ正負の方向に有意 (ボンフェローニ修正をして  $p < 0.05/4$ )。

### 3.4 結果

表1に結果の概要を示す。分析上の関心である依存文法および CCG に基づく予測子の結果のみ示している。なお、これら以外の予測子 (文節の位置、長さ、サブライザルに基づくもの) はいずれも予測された方向に有意であった。

CLT コストは、スピルオーバー領域においてのみ安定した正の効果があった。依存関係の距離に基づく効果はサンプル間で一貫しないものの、スピルオーバー領域で正の効果を持つ傾向にあった。文節基準と文字基準で結果に大きな違いは見られなかったが、文字基準のほうが係数が大きく  $p$  値が小さい傾向にあった。また、当該文節で完結する依存関係の数と、当該文節で適用される CCG 規則の数は、いずれも安定して負の効果があった。

## 4 議論

本分析の結果は、CLT コストが日本語においても読み時間を予測でき、依存関係の距離よりも局所性効果の予測子として優れていることを示している。この点で Isono (2024) の英語コーパスの分析と合致する結果となった。一方で、CLT コストの効果が観察されたのはスピルオーバー領域のみで、この点は英語コーパスにおいて CLT コストの効果が当該領域ですぐに観察されたのとは異なっている。これが言語の違いに由来するのか、コーパスや参加者の性質に由来するのかは直ちに明らかではないが、一つ考えられるのは、コーパスの性質の違いである。Isono (2024) で用いられていた Natural Stories コーパスは、統語的に複雑な文を含みつつ自然な文章になるように言語学者によってデザインされているのに対し、BCCWJ では実在のテキストをそのまま用いている。そのため、BCCWJ のほうが読みやすかったと考えられる。読み手は実験途中でもそれまでの刺激文の難易度に応じて読み方を変えるため (Isono & Hirose, 2023)、BCCWJ のほうがテンポよく読んでいて、統語的な複雑性を検知してボタン押しを思いとどまる前にいわば勢い余って次領域に進んでしまっている可能性がある。

依存関係に基づく距離もスピルオーバー領域で正の効果を見せたが、これは依存関係と CCG がそれぞれ別の処理過程 (たとえば意味と統語) において利用される表象を反映しているからだと考えられる (Isono, 2024)。

距離の指標に対応して、いわば距離を生み出す操作自体の数を数えているのが依存関係の数および CCG 規則適用の回数であるが、これらは一貫して負の効果を見せた。この結果も先行研究の観察と合致する (浅原ほか, 2019; Kajikawa et al., to appear; Konieczny & Döring, 2003; Vasishth & Lewis, 2006)。依存関係の数が持つ負の効果については、先行する依存部が多いほど主要部の予測がしやすくなるためだと説明されてきた (Konieczny & Döring, 2003; Levy, 2008; cf. Vasishth & Lewis, 2006)。CCG 規則適用の回数の効果も、同様に考えることができるかもしれない。たとえば、(4) の目的語関係節では、*attacked* が処理される際に、作業記憶にある  $S/(S \setminus NP)$  (埋め込み節主語) と  $NP/(S/NP)$  (先行詞 + *who*) との規則適用が順に発生する。主語しか文脈になければ次が自動詞である可能性も高いが、関係節であることでその可能性が減じられているので、論理的には関係節内のほうが動詞の予測はしやすく、その予測しやすさが規則適用の回数にも反映されていると考えうる。しかし、ここでの“予測”の意味には注

意しなければならない。そもそも CCG が文理解と親和的であるのは、CCG がある種の推論体系であり、文法内である程度の予測ができるからである。たとえば  $S/NP$  という範疇は次に  $NP$  が現れるという予測を表している。目的語関係節において一時的に作業記憶に 2 つの構成素が生じるのは、まさにこのような予測が文法内でできていないからであり、CLT はそれこそが作業記憶の負荷の原因であるとしているのだから、安易に文法外で解析器に予測を許すとモデル全体の仮定を崩しかねない。むしろ、ここでの予測とは、論理的推論ではなく範疇の活性化であると考えうる (cf. Huettig et al., 2022)。すなわち、 $NP/(S/NP)$  (先行詞 + *who*) という範疇が作業記憶にあると、そこに含まれる  $S/NP$  という範疇の表象が事前に活性化され、これが  $S/(S\backslash NP)$  (埋め込み節主語) と  $(S\backslash NP)/NP$  (埋め込み節動詞) を合成して  $S/NP$  を生じさせる際にその範疇へのアクセスを速める、というシナリオである。ただ、この仮説を厳密に述べるには、範疇についての長期記憶上の知識のモデルが不可欠であり、この点は将来の課題とする。

## 謝辞

BCCWJ-SPR2 のデータを提供くださった浅原正幸氏に感謝します。

## 参考文献

- Asahara, M. (2022). Reading time and vocabulary rating in the Japanese language: Large-scale Japanese reading time data collection using crowdsourcing. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 5178–5187.
- Asahara, M., Maekawa, K., Imada, M., Kato, S., & Konishi, H. (2014). Archiving and Analysing Techniques of the Ultra-Large-Scale Web-Based Corpus Project of NINJAL, Japan. *Alexandria*, 25(1-2), 129–148. <https://doi.org/https://doi.org/10.7227/ALX.0024>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bartek, B., Lewis, R. L., Vasishth, S., & Smith, M. R. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1178–1198. <https://doi.org/10.1037/a0024194>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. <https://arxiv.org/abs/1506.04967>
- Demberg, V. (2012). Incremental derivations in CCG. *Proceedings of the 11th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+11)*, 198–206.
- Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109, 192–210.
- Demberg, V., Keller, F., & Koller, A. (2013). Incremental, predictive parsing with psycholinguistically motivated Tree-Adjoining Grammar. *Computational Linguistics*, 39(4), 1025–1066. [https://doi.org/10.1162/COLLa\\_00160](https://doi.org/10.1162/COLLa_00160)
- Dotlačil, J. (2021). Parsing as a cue-based retrieval model. *Cognitive Science*, 45(e13020). <https://doi.org/10.1111/cogs.13020>
- Futrell, R., Gibson, E., & Levy, R. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science*, 44(3), e12814. <https://doi.org/10.1111/cogs.12814>
- Futrell, R., Gibson, E., Tily, H. J., Blank, I., Vishnevetsky, A., Piantadosi, S. T., & Fedorenko, E. (2021). The Natural Stories corpus: A reading-time corpus of English texts containing rare syntactic constructions. *Language Resources and Evaluation*, 55, 63–77. <https://doi.org/10.1007/s10579-020-09503-7>
- Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. P. Marantz, Y. Miyashita, & W. O’Neil (Eds.), *Image, language, brain* (pp. 95–126). MIT Press.
- Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, 29(2), 261–290. [https://doi.org/10.1207/s15516709cog0000\\_7](https://doi.org/10.1207/s15516709cog0000_7)
- Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. *Proceedings of NAACL 2001*. <https://doi.org/10.3115/1073336.1073357>
- Huettig, F., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition*, 224, 105050. <https://doi.org/https://doi.org/10.1016/j.cognition.2022.105050>

- Isono, S. (2024). Category Locality Theory: A unified account of locality effects in sentence comprehension. *Cognition*, 247, 105766. <https://doi.org/10.1016/j.cognition.2024.105766>
- Isono, S., & Hirose, Y. (2023). Pre-verb reactivation of arguments in sentence processing. *Glossa Psycholinguistics*, 2(1). <https://doi.org/10.5070/G6011180>
- Kajikawa, K., Yoshida, R., & Oseki, Y. (to appear). Dissociating syntactic operations via composition count. *Proceedings of CogSci 2024*.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580–602.
- Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, 29(6), 627–645. <https://doi.org/10.1023/A:1026528912821>
- Konieczny, L., & Döring, P. (2003). Anticipation of clause-final heads: Evidence from eye-tracking and SRNs. *Proceedings of the 4th International Conference on Cognitive Science*, 330–335.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. <https://doi.org/10.1016/j.cognition.2007.05.006>
- Levy, R., & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, 68(2), 199–222. <https://doi.org/10.1016/j.jml.2012.02.005>
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419. <https://doi.org/10.1207/s15516709cog0000.25>
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447–454. <https://doi.org/10.1016/j.tics.2006.08.007>
- Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koiso, H., Yamaguchi, M., Tanaka, M., & Den, Y. (2014). Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation*, 48, 345–371. <https://doi.org/10.1007/s10579-013-9261-0>
- Nakatani, K. (2023). Locality-based retrieval effects are dependent on dependency type: A case study of a negative polarity dependency in Japanese. In M. Koizumi (Ed.), *Issues in Japanese psycholinguistics from comparative perspectives (vol. 2)* (pp. 31–54). De Gruyter Mouton. <https://doi.org/10.1515/9783110778939-003>
- Nakatani, K., & Gibson, E. (2010). An on-line study of Japanese nesting complexity. *Cognitive Science*, 34(1), 94–112. <https://doi.org/10.1111/j.1551-6709.2009.01067.x>
- Ono, H., & Nakatani, K. (2014). Integration costs in the processing of Japanese *wh*-interrogative sentences. *Studies in Language Sciences*, 13, 13–31.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners.
- Safavi, M. S., Husain, S., & Vasishth, S. (2016). Dependency resolution difficulty increases with distance in Persian separable complex predicates: Evidence for expectation and memory-based accounts. *Frontiers in Psychology*, 7(403). <https://doi.org/10.3389/fpsyg.2016.00403>
- Shain, C., & Schuler, W. (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2679–2689. <https://doi.org/10.18653/v1/D18-1288>
- Shain, C., Van Schijndel, M., Futrell, R., Gibson, E., & Schuler, W. (2016). Memory access during incremental sentence processing causes reading time latency. *Proceedings of the workshop on computational linguistics for linguistic complexity*, 49–58.
- Stanojević, M., Brennan, J. R., Dunagan, D., Steedman, M., & Hale, J. T. (2023). Modeling structure-building in the brain with CCG parsing and large language models. *Cognitive Science*, 47(7), e13312. <https://doi.org/10.1111/cogs.13312>
- Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, 116(1), 71–86. <https://doi.org/10.1016/j.cognition.2010.04.002>
- Steedman, M. (2000). *The syntactic process*. MIT Press. <https://doi.org/10.7551/mitpress/6591.001.0001>
- Steedman, M. (2023). On internal merge. *Linguistic Inquiry, Advance publication*. [https://doi.org/10.1162/ling\\_a.00521](https://doi.org/10.1162/ling_a.00521)
- Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, 82(4), 767–794. <https://doi.org/10.1353/lan.2006.0236>
- Yoshikawa, M., Noji, H., & Matsumoto, Y. (2017). A\* CCG parsing with a supertag and dependency factored model. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 277–287. <https://doi.org/10.18653/v1/P17-1026>
- 戸次 大介. (2010). *日本語文法の形式理論*. くろしお出版.
- 梶川 康平, 吉田 遼, 大関 洋平. (2023). CCGによる日本語文処理のモデリング. In *言語処理学会第29回年次大会*.
- 浅原 正幸, 小野 創, 宮本 エジソン 正. (2019). BCCWJ-EyeTrack. *言語研究*, 156, 67–96. <https://doi.org/10.11435/gengo.156.0.67>