

CCG 統語解析器 lightblue と定理証明器 wani による JSeM Verbs データセットの自動推論

松原舞 富田朝 戸次大介
お茶の水女子大学

{matsubara.mai, tomita.asa, bekki}@is.ocha.ac.jp

概要

自然言語理解の研究において、大規模言語モデル (LLM) とは対照的なアプローチとして、合理主義自然科学のパラダイムに基づく言語学的パイプラインの研究が進展している。これは理論言語学に基づく統語解析、意味合成、自動証明のモジュールを接続したものであり、仮説検証を通じてモデルを改善する点で、経験主義的な LLM と相補的な関係にある。本論文では、CCG に基づく統語解析器 lightblue と、DTS のための定理証明器 wani を組み合わせた自然言語推論システムを用いて、JSeM 日本語テストセットの Verb セクションの推論問題を解くことを試みた。この試みは、システムの評価、エラー分析、理論へのフィードバックを含む、合理主義的計算言語学の実践例の一つである。

1 はじめに

近年の研究において、大規模言語モデル (LLM) の振る舞いが、人間の言語理解のモデルとは異なっている点が複数指摘されている。例として、LLM による推論は人間の推論が持つ体系性 (systematicity) を欠くという指摘 [1] や、LLM が複雑な問題を解けるかどうかはその問題の出現頻度に依存している、という指摘 [2] などが挙げられる。

一方、計算言語学においては、理論言語学が仮定する人間の言語理解モデルに従って統語解析、意味合成、自動証明のモジュールを接続した言語学的パイプライン [3, 4, 5, 6, 7] による自然言語理解の研究が進んでいる。言語学的パイプラインは、仮説検証を繰り返すことでモデルを改善する点で古典的な合理主義自然科学のパラダイムを踏襲しており、経験主義を標榜する大規模言語モデルとは相補的な関係にある。

言語学的パイプラインの近年の進展には、複数の要因が貢献している。組合せ範疇文法 (Combinatory Categorical Grammar, 以下 CCG) [8] の成熟により、高速かつ頑健な CCG 統語解析器 [9, 10, 11] が実現したこと、また依存型理論 (Dependent Type Theory, 以下 DTT) [12] に基づく証明論の意味論 [13, 14, 15, 16] が登場し、形式意味論による意味の記述と定理自動証明器の実装を組み合わせたことが可能となったこと等が挙げられる。

本稿ではそのような試みのひとつである CCG 統語解析器 lightblue [17] と DTT 定理証明器 wani [18, 19] を組み合わせた自然言語推論システムによって、JSeM 日本語テストセット [20, 21] (§3.5 節参照) の Verb セクションの自然言語推論問題を解くことを試みる。この過程には、推論システムの評価、エラー分析、そしてパイプライン上の各理論へのフィードバックが含まれており、合理主義的計算言語学の一つの実践例として紹介する。

2 先行研究

自然言語推論 (Natural Language Inference, 以下 NLI) は、前提と仮説の間の含意関係の有無を判定するタスクであり、多くの意味処理タスクの基盤であると考えられている。NLI に対するアプローチとしては LLM によるもの [22, 23] が主流であるが、計算言語学においても §1 節で挙げたさまざまなアプローチが存在する。

Bos [3] による Boxer は、言語学的パイプラインのさきがけである。CCG 統語解析器の出力をもとに、談話表現理論 (Discourse Representation Theory, DRT) [24] に基づいた意味表示を自動で生成する。得られた意味表示を一階述語論理の論理式に変換することで、標準的な定理証明システムを用いた推論が可能である。

Abzianidze [5] では、CCG 統語解析器から得られた統語構造をもとにラムダ論理形式 (LLF) を生成し、それをもとに Natural Tableaux [25] を用いて推論を行う一連の自動推論のパイプラインが提案された。しかし、Bos [3], Abzianidze [5] の共通の課題として、複雑な言語現象を正しく扱うことができない。これは自動証明可能な論理体系を採用していることのトレードオフである。

この課題に対する一つの解決策は、高階論理を用いることである。Chatzikyriakidis and Luo [4] では現代的型理論 (Modern Type Theory, MTT) [26] に基づく MTT 意味論 [15] をもとに意味表示を構成し、証明支援系 Coq [27] を用いた半自動推論が行われた。しかし、統語解析器との接続は行われておらず、自然言語から意味表示への変換は手動である。

この点を改良したのは Mineshima et al. [6] による `ccg2lambda`¹⁾ である。`ccg2lambda` は、CCG 統語解析器 Jigg [28] および `depccg` [11] と、証明支援系 Coq [27] を組み合わせることで、英語テキストの含意関係を自動で判定するシステムである。また、Mineshima et al. [7] では `ccg2lambda` の日本語対応が実現している。

Chatzikyriakidis and Luo [4] と Mineshima et al. [6] はともに高階論理 (Higher-Order Logic, HOL) を採用しており、一般化量詞のような一階述語論理では表現できない言語現象も扱うことができる。その反面、これらのシステムにおいては一部の単純な証明は自動化している一方で、複雑な証明の多くの部分は手動で行われるか [4], もしくはタイムアウトとして扱われている [6]。

3 システムの概要

本研究で使用した自動推論システム²⁾もまた、言語学的パイプラインによって自動推論を実現している。本システムの構想は [29] において構想が示されたもので、統語解析には CCG 統語解析器 `lightblue` [17, 30] を、意味の理論には依存型意味論 (Dependent Type Semantics, DTS) [31] を採用し、定理証明には DTT のための自動証明器 `wani` [18, 19] を用いる。推論システムのパイプラインを図 1 に示す。前提 T_1, \dots, T_n と仮説 H に対して、一文ごとに統語解析・意味合成と型検査を繰り返し、全ての文で型

検査が成功したときに証明探索を行う。各文のための型検査においては、それまでの文の意味表示が、文脈に追加されていくのが重要な点である。このことによって、DTS の理論に従い、型検査の過程の中で照応解決を行うことが可能となっている。

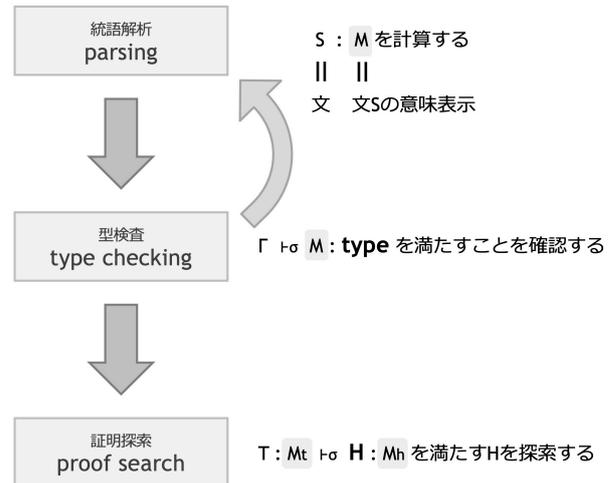


図 1 lightblue と wani による自動推論

本節では、まず `lightblue` と `wani` のベースとなっている DTS (§ 3.1 節) について紹介し、続いて推論の各工程 (統語解析 (§ 3.2 節), 型検査 (§ 3.3 節), 証明探索 (§ 3.4 節)) について詳細を述べ、最後に評価に用いる日本語意味論テストセット JSeM (§ 3.5 節) を紹介する。

3.1 依存型意味論 (DTS)

DTS は、DTT に基づいた自然言語のための証明論的意味論である。DTS は項に依存した型を記述するため、 S_1, \dots, S_n において S_n の意味を記述する際には、 S_1, \dots, S_{n-1} の意味に依存した文の意味を記述することが可能である。この特徴から前提束縛や照応解決などの複雑な言語現象を扱うことができる。DTS を含む型理論では、判定 $\Gamma \vdash M : A$ は文脈 Γ において項 M の型が A であることを表す。カーリー＝Howard 対応 (図 2) より、項は証明に、型は命題に対応することから、証明論においては、前述の判定 $\Gamma \vdash M : A$ は前提 Γ のもとで命題 A の証明 M が存在することを表す。このことから、DTS において前提と仮説の間の含意関係を推論することは、前提 Γ のもとで命題 A の証明項 M を探索することに帰結する。

1) <https://github.com/mynlp/ccg2lambda>
 2) <https://github.com/DaisukeBekki/lightblue>

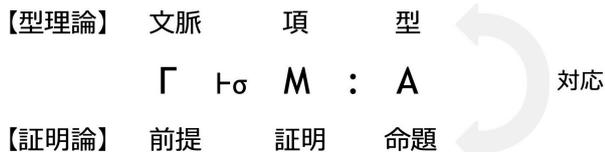


図2 カリー=ハワード対応

3.2 統語解析 (syntactic parsing)

CCG 統語解析器 lightblue³⁾ [17] を用いることにより統語解析を行う。lightblue は日本語の文を日本語 CCG[30] に基づいてパースし、統語構造および DTS における意味合成の過程 (のリスト) を出力する。lightblue が出力する意味表示は DTS の型と対応している。

3.3 型検査 (type checking)

型検査とは統語解析の結果の妥当性を確認するための操作である。その過程で、照応や前提の解決に必要な文脈が正しく計算されるという点が、他の意味の理論に対する DTS の優位性である。DTS には、文の意味表示の型は type でなければいけないという制約があり、これを満たしているかどうかを型検査によって確認する。また $S_n (n \geq 2)$ においては、DTS では § 3.1 節で述べたように、 S_1, \dots, S_{n-1} の意味を加味した分析ができることから、文脈 Γ に S_1, \dots, S_{n-1} の意味を与えた上で検査を行う。また、項の中に未指定項 (@項) を含む場合は、それを具体的な項に置き換える照応解決の役割も持つ。

本研究で使用した推論システムでは、統語構造の曖昧性、照応解決の曖昧性を加味した上で、統語解析および型検査を自動で行う。

3.4 証明探索 (proof search)

定理証明器 wani [18, 19] を用いて証明探索を行う。wani は DTS の断片⁴⁾ についての自動定理証明器であり、DTS 形式の前提と仮説を受け証明が可能であった場合は証明項を含む結果を出力する。本研究で使用した推論システムは、前提 T_1, \dots, T_n に対して、仮説 H についての証明項が返された場合は含意 (yes)、仮説 $\neg H$ についての証明項が返された場合は矛盾 (no)、いずれも証明項が返されなかった場合はどちらでもない (unknown) と判定する。

3) <https://github.com/DaisukeBekki/lightblue>

4) Π 型, Σ 型, 枚挙型 (enumeration type) が使用できる。

3.5 JSeM : 日本語意味論テストセット

JSeM⁵⁾ [20, 21] は日本語の意味的な言語現象に基づく含意関係のデータセットで、英語用のテストセット FraCaS [32] の日本語訳にあたる問題や、日本語独自の言語現象に対応した問題からなる。各問題は前提、仮説、判定 (正解ラベル) などで構成されており、Generalized Quantifier (一般量化子)、Verbs (動詞) など、言語現象ごとにセクションが分けられている。判定には yes, no, unknown, undef (与えられた情報のみからは判断ができない) の4つのラベルのいずれかが当てられている。このラベルと含意関係の推論結果の一致率 (正答率) によって、含意関係認識システムを評価することができる。

4 評価

本研究では JSeM のうち Verbs (動詞) セクション⁶⁾ の推論に取り組んだ。

4.1 結果とエラーの分析

Verbs セクションは全部で 36 題あり、その全てで型検査に成功した。また 20 題 (55.5%) で推論が正解した。これは ccg2lambda と同等のスコアであった。以下にそれぞれのスコアを掲載する (表 1, 表 2)。統語解析および型検査では、上位 5 位までを評価対象 (n-best = 5) とし、証明探索では、探索の最大の深さ (maxDepth) は 5 と設定した。各工程の結果をもとに分析した結果、1 つの問題に対して複数のエラーが発生しているものも見つかった。以下では推論が不正解だったもの (正解ラベルが Yes に対して Unk と推論されたもの) のうち 2 種類のエラーについてそれぞれ分析を行う。

4.1.1 複合動詞に由来するエラー (JSeM ID 693)

【前提】 ITEL は 1993 年に MTALK を作った。

【仮説】 ITEL は 1993 年に MTALK を作り終えた。

【正解ラベル】 yes

仮説について、「MTALK を作り」+「終えた」と解析された。この場合、「MTALK を作り」が連用節として分析されることで、「終えた」の項とし

5) <https://github.com/DaisukeBekki/JSeM>

6) <https://github.com/DaisukeBekki/JSeM/blob/master/data/v1.0/Verbs.xml>

表1 本システムの推論結果

		正解ラベル			
		Yes	No	Unk	Other
推論結果	Yes	13	0	0	0
	No	0	0	0	0
	Unk	16	0	7	0
	Other	0	0	0	0

表2 ccg2lambda の推論結果

		正解ラベル			
		Yes	No	Unk	Other
推論結果	Yes	20	0	0	0
	No	0	0	0	0
	Unk	1	0	0	0
	Other	8	0	7	0

て「MTALK を」というヲ格名詞を与えることができず、「終えた」の項として pro (空名詞句) が渡される。この解析は文法的には誤りではないが、「MTALK を」+「作り終えた」という統語構造がより自然である。後者の構造が導出されない理由は、lightblue の複合動詞解析モジュールが未完成であるためである。

4.1.2 知識に由来するエラー (JSeM ID 727)

【前提】 太郎が障子を破いた。

【仮説】 障子が破れた。

【正解ラベル】 yes

(任意の x, y について) 「 x が y を破いた」ならば「 y が破れた」が成立するが、lightblue 上では「破く」と「破れる」は別の動詞であるため、この二つを結びつけるには(動詞の自他に関する)ある種の公理を導入する必要がある。

4.2 エラーの改善に向けて

複合動詞に由来するエラーについては、すでに研究が進められている [33]。そこで、本稿では知識に由来するエラーについて改善案を述べる。

寺村 [34] によると、「破く」-「破れる」のような動詞の自他の形態的な対立関係は 10 種類に分類できるとされている。分類ごとに系統的な振る舞いをみせるため、各動詞がどの分類に属しているかの情報が得られれば、推論に必要な公理を必要に応じて

提供する機能を実装することは難しくない。しかしながら、lightblue の辞書は少数の open words と多数の closed words から構成されており、前者は日本語 CCG [30] から取り込んだ手書きの語彙項目、後者は JUMAN⁷⁾ の辞書データから自動変換した語彙項目である。「破く」「破れる」のような動詞の語幹は後者に属しているが、それらのうち、どの動詞にどの分類情報を与えるかは自明ではなく、今後の研究課題の一つとしたい。

一方で、以下のような問題では「ツバメは鳥である」という世界知識が必要となるが、このような知識は膨大であり辞書に逐次与えるのは困難である。

【前提】 ツバメが空を飛んだ。

【仮説】 鳥が空を飛んだ。

そこで、このような包含関係に関する世界知識が必要になった際には、LLM を活用して公理を生成することも検討している。

5 おわりに

本稿では、lightblue と wani を組み合わせて日本語の推論テストセットを対象とした自動推論を行い、その評価と分析からフィードバックに至る過程を紹介した。

なお、統語構造や照応解決には曖昧性があることから、上位 5 位までを評価対象とし、得られた推論結果のうち、ひとつでも正解ラベルと一致していれば推論が成功したものとして評価を行った。型検査では 3 つ目の結果が適切だった例が見つかったため、統語解析結果の順位が低い意味表示の扱いをどうすべきか、という(曖昧性のない統語解析器を用いていた研究では生じなかった)新たな理論面・実装面にまたがる問題も生じている。同時に、推論システムの評価方法を再検討する必要も生じている。

今後は上述の課題や、知識を補完する手法を検討しつつ、JSeM の他のセクションにも取り組み、推論結果の分析を行うことで、本システムの人間の言語理解の再現へ向けたさらなる発展に貢献したい。

7) <https://nlp.ist.i.kyoto-u.ac.jp/?JUMAN>

謝辞 本研究の一部は, JST CREST JPMJCR20D2, および JSPS 科研費 JP23H03452 の助成を受けたものである。

参考文献

- [1] 谷中瞳, 峯島宏次. AI は言語の基盤を獲得するか: 推論の体系的観点から. 認知科学, 第 31 巻, 2024.
- [2] R. Thomas McCoy, Shunyu Yao, Dan Friedman, Mathew D. Hardy, and Thomas L. Griffiths. Embers of autoregression show how large language models are shaped by the problem they are trained to solve. **Proceedings of the National Academy of Sciences**, Vol. 121, No. 41, p. e2322420121, 2024.
- [3] Johan Bos. Wide-coverage semantic analysis with boxer. In **Semantics in Text Processing, STEP 2008 Conference Proceedings**, pp. 277–286, 2008.
- [4] Stergios Chatzikyriakidis and Zhaohui Luo. Natural language inference in coq. **Journal of Logic, Language and Information**, Vol. 23, pp. 441–480, 2014.
- [5] Lasha Abzianidze. A tableau prover for natural logic and language. In **EMNLP2015**, pp. 2492–2502, 2015.
- [6] Koji Mineshima, Pascual Martínez-Gómez, Yusuke Miyao, and Daisuke Bekki. Higher-order logical inference with compositional semantics. In **Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing**, pp. 2055–2061, 2015.
- [7] Pascual Martínez-Gómez Yusuke Miyao Koji Mineshima, Ribeka Tanaka and Daisuke Bekki. Building compositional semantics and higher-order inference system for a wide-coverage japanese ccg parser. In **Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing**, pp. 2236–2242, 2016.
- [8] Mark Steedman. **The Syntactic Process**. MIT Press, 2000.
- [9] S. Clark and J. R. Curran. Widecoverage efficient statistical parsing with ccg and log-linear models. **Computational Linguistics**, Vol. 33, No. 4, pp. 493–552, 2007.
- [10] Mike Lewis and Mark Steedman. A* ccg parsing with a supertag-factored model. In **the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 990–1000. Association of Computational Linguistics, 2014.
- [11] Masashi Yoshikawa, Hiroshi Noji, and Yuji Matsumoto. A* ccg parsing with a supertag and dependency factored model. In **the 55th Annual Meeting of the Association for Computational Linguistics**, pp. 277–287, 2017.
- [12] Per Martin-Löf and Giovanni Sambin. **Intuitionistic type theory**, Vol. 17. Bibliopolis, 1984.
- [13] Göran Sundholm. **Proof theory and meaning**, Vol. III, pp. 471–506. Kluwer, Reidel, 1986.
- [14] Aarne Ranta. **Type-Theoretical Grammar**. Oxford University Press, 1994.
- [15] Zhaohui Luo. Formal semantics in modern type theories with coercive subtyping. **Linguistics and Philosophy**, Vol. 35, No. 6, 2012.
- [16] Daisuke Bekki. Representing anaphora with dependent types. In Nicholas Asher and S. V. Soloviev, editors, **Logical Aspects of Computational Linguistics (8th international conference, LACL2014, Toulouse, France, June 2014 Proceedings)**, LNCS 8535, pp. 14–29. Springer, Heiderburg, 2014.
- [17] Daisuke Bekki and Ai Kawazoe. Implementing variable vectors in a ccg parser. In Christian Retoré and Sylvain Pogodalla, editors, **Logical Aspects of Computational Linguistics (9th international conference, LACL2016, Nancy, France, December 2016 Proceedings)**, pp. 52–67. Springer, Heiderburg, 2016.
- [18] Hinari Daido and Daisuke Bekki. Development of an automated theorem prover for the fragment of dts. In **the 17th International Workshop on Logic and Engineering of Natural Language Semantics (LENLS17)**, 2017.
- [19] 大洞日音. DTS の部分体系を用いた定理自動証明器への等号型の導入. Master's thesis, お茶の水女子大学, 2022.
- [20] Ai Kawazoe, Ribeka Tanaka, Koji Mineshima, and Daisuke Bekki. A framework for constructing multilingual inference problem sets: Highlighting similarities and differences in semantic phenomena between english and japanese. In **1st International Workshop on the Use of Multilingual Language Resources in Knowledge Representation Systems (MLKRep2015)**, 2015.
- [21] Ai Kawazoe, Ribeka Tanaka, Koji Mineshima, and Daisuke Bekki. An inference problem set for evaluating semantic theories and semantic processing systems for japanese. In **the 17th International Workshop on Logic and Engineering of Natural Language Semantics (LENLS12)**, pp. 67–73, 2015.
- [22] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. Albert: A lite bert for self-supervised learning of language representations. In **International Conference on Learning Representations (ICLR 2020)**, 2020.
- [23] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. **The Journal of Machine Learning Research**, Vol. 21, No. 140, pp. 5485–5551, 2020.
- [24] Hans Kamp. A theory of truth and semantic representation. In **Formal Methods in the Study of Language**, pp. 277–322, 1981.
- [25] Reinhard Muskens. An analytic tableau system for natural logic. In **Logic, Language and Meaning**, pp. 104–113. Springer, 2010.
- [26] Luo and Zhaohui. **Computation and Reasoning: A Type Theory for Computer Science**. Clarendon Press, 1994.
- [27] Yves Bertot and Pierre Castéran. Interactive Theorem Proving and Program Development Coq' Art: The Calculus of Inductive Constructions. Springer, 2004.
- [28] Hiroshi Noji and Yusuke Miyao. Jigg: A framework for an easy natural language processing pipeline. In **the 54th Annual Meeting of the Association for Computational Linguistics**, pp. 103–108, 2016.
- [29] 戸次大介. 依存型意味論による自然言語の意味の理論: 見通しと課題. 第 37 回人工知能学会全国大会論文集, 2023.
- [30] 戸次大介. 日本語文法の形式理論 – 活用体系・統語構造・意味合成 –. くろしお出版, 2010.
- [31] Daisuke Bekki and Koji Mineshima. Context-passing and underspecification in dependent type semantics. In **Modern Perspectives in Type Theoretical Semantics**, pp. 11–41. Springer, 2017.
- [32] Robin Cooper, Dick Crouch, Jan van Eijck, Chris Fox, Josef van Genabith, Jan Jaspars, Hans Kamp, David Milward, Manfred Pinkal, Massimo Poesio, and Steve Pulman. A strategy for building a framework for computational semantics(the way forward). In **FraCaS: A Framework for Computational Semantics**, 1996.
- [33] Asa Tomita, Hitomi Yanaka, and Daisuke Bekki. Reforging : A method for constructing a linguistically valid Japanese CCG treebank. In Neele Falk, Sara Papi, and Mike Zhang, editors, **Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop**, pp. 196–207, St. Julian's, Malta, 2024. Association for Computational Linguistics.
- [34] 寺村秀夫. 日本語のシンタクスと意味 I. くろしお出版, 1982.

付録

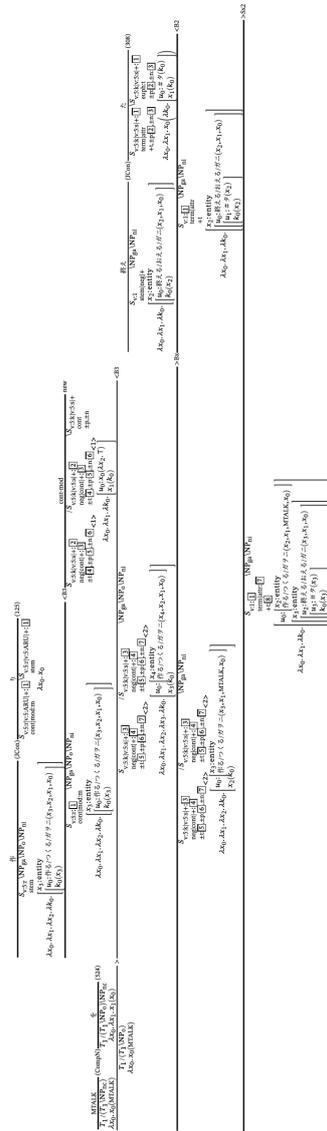


図 3 JSeM ID 693 の仮説の統語構造の一部

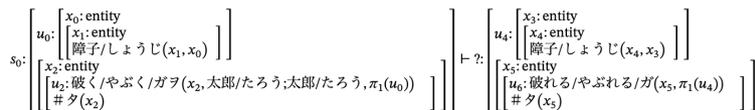


図 4 JSeM ID 727 の証明探索