

# 代表性をもつ日本語 CCG ツリーバンクの自動構築と評価

富田 朝 (指導教員: 戸次 大介)

## 1 はじめに

自然言語処理における推論タスクは、近年大きな進展を遂げている。[4] 大規模言語モデル (LLM) の精度向上とともに、言語学的アプローチに基づく自動推論の研究も発展している。特に、依存型意味論 [2] や自動証明器 Wani[5] の登場により、理論言語学に基づく言語処理の可能性が拡大している。このアプローチでは、統語論として組合せ範疇文法 (CCG) [9, 10]、意味論として依存型意味論を採用し、これらの分析を推論プロセスに統合することで、推論結果とその過程を証明図として提示し、結果の妥当性保証や誤りの原因特定が可能となる。

推論の精度は統語解析や意味解析の出力に大きく依存する。しかし現在の開発されている日本語 CCG 統語解析器 [14] は、評価データセットで高精度を示す一方、受身や使役などの複雑な構文に対しては誤りが含まれることが知られている。これは学習・評価に用いられるデータセット自体に誤りが含まれる [3] ことが一因である。

統語解析器の学習・評価には、統語ラベル付きコーパスであるツリーバンクが使用される。様々な言語や形式文法に基づくツリーバンクが開発され [6, 13]、その語彙網羅性や言語処理タスクでの有用性が評価されているが、これらツリーバンクの言語学的妥当性に関する評価は十分に行われていないのが現状である。

本研究では、統語論と意味論に基づく新たなツリーバンク lightblue CCGbank を構築し、その言語学的妥当性を評価するための評価手法を提案する。そして lightblue CCGbank を対象に言語学的妥当性の評価実験を行う。

## 2 日本語 CCG 統語解析器 lightblue

lightblue[1] は CCG に基づく統語構造と依存型意味論に基づく意味表示を並行して計算することのできる解析器である。CCG ツリーバンクを学習データに用いたニューラル統語解析器 [14] とは異なり、あらかじめ形態素解析器 Juman[7] の辞書データから変換した語彙を含む CCG の辞書と、CCG の組合せ規則を用いて統語・意味解析を行うため、学習データとしてツリーバンクを必要としないという特徴がある。さらに、lightblue の辞書には、用言の活用形や活用の種類、テンスをはじめとした、詳細な統語素性の情報を含むこと、CCG 統語構造と並行して、依存型意味論に基づく意味表示を出力できること、などの利点もある。

## 3 言語学的妥当なツリーバンクの構築

lightblue は上記の利点がある一方で、Juman から変換した辞書に用言の項構造の誤りが含まれるという課題を残している。そこで、lightblue を用いた解析時に、他の言語資源 [8, 12] から用言の項構造を抽出し、lightblue の語彙項目を削除・追加する項構造合成モジュールを作成した [11]。このシステムの概要図を図 1 に示す。

この項構造合成モジュールと、解析器 lightblue を用いたツリーバンクの構築手法を Reforging と呼び、それによって言語学的に妥当な日本語 CCG ツリーバンク lightblue CCGbank<sup>1</sup>の構築を行った。lightblue CCGbank は、ABC ツリーバンク [8] から抽出した 13,653 文に、CCG 統語構造および依存型意味論に基づく意味表示を付与したデータセットである。

## 4 ツリーバンクの妥当性の評価

一般に、ツリーバンクの評価には、辞書のエントリ数、辞書被覆率、パーザの解析精度などの指標が用いられるが、評価指標はいずれも妥当性の評価としては完全なものではない。そこで、本研究では、妥当性の評価手法を新たに提案する。

### 4.1 統語論に基づく評価

lightblue CCGbank は、ABC ツリーバンクから抽出された文を基にしており、それぞれの統語構造には対応する ABC ツリーバンクの統語構造が存在する。この関係を利用して、ABC ツリーバンクの統語構造が妥当であると仮定し、両者の一致度をスコア化することで、lightblue CCGbank の統語構造の信頼性を評価する。ABC ツリーバンクで使用される ABC 文法は、関数適用規則に関数合成規則を加えた範疇文法であり、CCG との比較には注意が必要である。具体的には、ABC 文法と CCG では統語範疇の定義や unary 規則の定義が異なるため、直接的な比較は不可能である。このため、ABC ツリーバンクの統語範疇を CCG で使用される統語範疇に変換し、一致度をスコア化することで評価を行う。この評価手法の利点には、CCG の空範疇と ABC ツリーバンクの unary 規則の比較が可能であること、用言の分析における差異に対応できることが挙げられる。しかし、ABC ツリーバンクの全ての統語構造が正しいとは限らないため、ABC ツリーバンクを正解データとして仮定する必要がある点や、ABC ツリーバンクに含まれない統語素性の評価ができないという課題も残されている。

### 4.2 意味論に基づく評価

lightblue CCGbank に含まれるすべての統語構造には、依存型意味論に基づく意味表示が付与されている。この意味表示を活用し、型理論の観点からその妥当性を評価する手法を提案する。まず、lightblue を用いて型検査を行い、型検査に通過する意味表示の割合を計算する。型検査とは、依存型意味論の意味表示が整合的な型を持つかを判定する手続きで、意味表示が type という型を持つことを証明できれば、型検査は成功し、統語構造と意味合成の整合性を担保できる。また、統語構造に対して、意味表示のレベルで型理論に基づいた評価を実施できる点も本手法の強みである。一方で、型検査を通過した意味表示と対になる統語構造が必ずしも妥当であるとは限らないため、この評価指標のみ

<sup>1</sup><https://github.com/morning85/lightblueCCGbank> に公開している

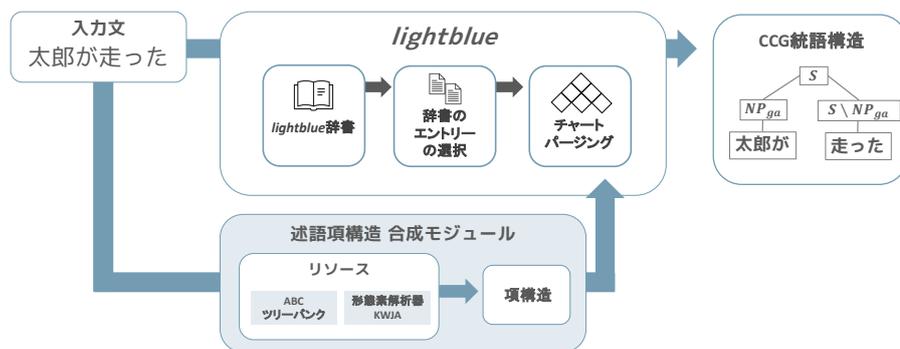


図 1: Reforging 概要図

で言語的妥当性を完全に保証することは難しい。より厳密な評価を行うには、前節の評価手法をはじめ、他の妥当性の評価指標と組み合わせることが望ましい。

### 4.3 評価実験

lightblue CCGbank に含まれる各ジャンルからランダムにサンプリングを行い、合計で 70 文を対象に評価を実施した。統語的な評価としては、統語論に基づく評価方法に従い、統語構造を 100 点満点で評価し、ジャンルごとの平均スコアを算出した。意味論的な評価としては、型検査に成功したデータの割合を計測した。また、総合的な評価として、統語構造のスコアが 50 点以上であり、かつ型検査に通過しているデータの割合を計算した。

### 4.4 実験結果

表 1: 言語学的妥当性の評価結果

ジャンル	データ数	スコア平均	型検査通過率	総合
青空文庫	75	43.1	69.3	21.1
聖書	24	37.8	58.3	17.2
書籍	6	43.0	66.7	25.0
辞書	60	53.0	48.3	25.9
会議録	21	32.9	90.5	16.0
フィクション	18	49.2	77.8	33.3
法律	6	12.0	66.7	0.0
その他	30	49.5	73.3	30.2
ニュース	30	40.7	76.7	11.8
ノンフィクション	6	40.0	100.0	25.0
話し言葉	30	31.8	73.3	16.7
テッドトーク	15	52.9	73.3	28.6
教科書	120	48.1	55.0	23.1
ウィキペディア	15	37.7	66.7	16.7
Total	456	40.8	64.9	22.3

実験の結果を表 1 に示す。

全体の平均スコアは 40.8 点であり、評価対象の文 456 文中 306 文が型検査に通過し、通過率は 64.9% であった。

スコア平均が最も高かったジャンルは辞書で、53.0% の統語構造が他コーパスと一致した。また、ノンフィクションは 6 文中全ての文が型検査を通過し、14 ジャンルで最も高い通過率となった。総合的な評価が最も高かったフィクションでは、3 割のデータで妥当性が保証された一方、法律、ニュース、聖書などのジャンルでは総合値が伸び悩んだ。この結果は、lightblue がドメイン固有の表現が含まれる文の解析に課題があるということを示唆しており、lightblue の辞書の拡張に

よって妥当性のさらなる向上が見込まれる。

## 5 おわりに

本研究では、統語論および意味論に基づく評価指標を提案し、ツリーバンクの妥当性を自動的に多角的に評価する枠組みを確立した。この手法により、従来の評価では捉えきれなかった統語構造の妥当性を詳細に分析することが可能となった。ツリーバンクの整備は、推論精度や説明可能性の向上に寄与し、今後は他のデータセットへの適用や評価手法の精度向上、フィードバック機構の導入を検討する予定である。

## 参考文献

- [1] Bekki, D. and Kawazoe, A.: Implementing Variable Vectors in a CCG Parser, in *Logical Aspects of Computational Linguistics. Celebrating 20 Years of LACL (1996–2016)*, pp. 52–67, Berlin, Heidelberg (2016), Springer Berlin Heidelberg.
- [2] Bekki, D. and Mineshima, K.: *Context-Passing and Underspecification in Dependent Type Semantics*, pp. 11–41, Springer International Publishing, Cham (2017).
- [3] Bekki, D. and Yanaka, H.: Is Japanese CCGBank empirically correct? A case study of passive and causative constructions, in *Proceedings of the 21st International Workshop on Treebanks and Linguistic Theories (TLT, GURT/SyntaxFest 2023)*, pp. 32–36, Washington, D.C. (2023), Association for Computational Linguistics.
- [4] Cobbe, K., Kosaraju, V., Bavarian, M., Chen, M., Jun, H., Kaiser, L., Plappert, M., Tworek, J., Hilton, J., Nakano, R., Hesse, C. and Schulman, J.: Training Verifiers to Solve Math Word Problems, *CoRR*, Vol. abs/2110.14168, (2021).
- [5] Daido, H. and Bekki, D.: 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)* (2020).
- [6] Hockenmaier, J. and Steedman, M.: CCGbank: A Corpus of CCG Derivations and Dependency Structures Extracted from the Penn Treebank, *Computational Linguistics*, Vol. 33, No. 3, pp. 355–396 (2007).
- [7] Kawahara, D. and Kurohashi, S.: Case Frame Compilation from the Web using High-Performance Computing, in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy (2006), European Language Resources Association (ELRA).
- [8] Kubota, Y., Mineshima, K., Hayashi, N. and Okano, S.: Development of a General-Purpose CCG Grammar Treebank, in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pp. 5195–5201, Marseille, France (2020), European Language Resources Association.
- [9] Steedman, M.: *Surface Structure and Interpretation*, The MIT Press, Cambridge (1996).
- [10] Steedman, M.: *The Syntactic Process*, MIT Press (2000).
- [11] Tomita, A., Yanaka, H. and Bekki, D.: Reforging : A Method for Constructing a Linguistically Valid Japanese CCG Treebank, in Falk, N., Papi, S. and Zhang, M. eds., *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.
- [12] Ueda, N., Omura, K., Kodama, T., Kiyomaru, H., Murawaki, Y., Kawahara, D. and Kurohashi, S.: KWJA: A Unified Japanese Analyzer Based on Foundation Models, in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Toronto, Canada (2023).
- [13] Uematsu, S., Matsuzaki, T., Hanaoka, H., Miyao, Y. and Mima, H.: Integrating Multiple Dependency Corpora for Inducing Wide-coverage Japanese CCG Resources, in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1042–1051, Sofia, Bulgaria (2013), Association for Computational Linguistics.
- [14] Yoshikawa, M., Noji, H. and Matsumoto, Y.: A\* CCG Parsing with a Supertag and Dependency Factored Model, in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 277–287, Vancouver, Canada (2017), Association for Computational Linguistics.