

日本語CCGツリーバンクの言語学的妥当性の評価

富田朝¹ 谷中瞳² 戸次大介¹

お茶の水女子大学¹ 東京大学²
tomita.asa@is.ocha.ac.jp



Ochanomizu University



THE UNIVERSITY OF TOKYO

はじめに：コーパスと評価

コーパス：新聞記事などのテキストや音声データをデータセットしてまとめた言語資源

- ・コーパス自体の質や妥当性を評価するのは難しい
- ・特に、専門的な知識が必要となる言語資源は人手の評価が一般的で、自動評価する方法はあまり確立されていない
- ・データ数の多いCCGツリーバンクの「妥当性」を自動で評価するための手法を2つ提案する

背景：なぜ妥当なツリーバンクが必要なのか

CCGツリーバンク：CCGの統語構造のコーパス

問題点

誤りを含むツリーバンクで学習された統語解析器は、
誤りを含む統語構造を出力する
→ 誤った推論へとつながる



図1：形式文法に基づいた推論パイプライン

LLMでの推論ではダメなのか？

LLM：なぜ誤りが生じたのかの原因の特定が難しい

形式文法：誤りの原因の特定が可能

→ 信頼性につながる

ツリーバンクの構築 (Tomita et al., EACL 2024)

- 日本語CCG統語解析器lightblue (Bekki and Kawazoe 2016) を用いて統語構造の自動アノテーションを行う

問題点

lightblueが参照する辞書に、述語項構造に関する誤りが含まれている

述語項構造 合成モジュールを開発し、他の言語資源から項構造情報を抽出し合成する手法を提案

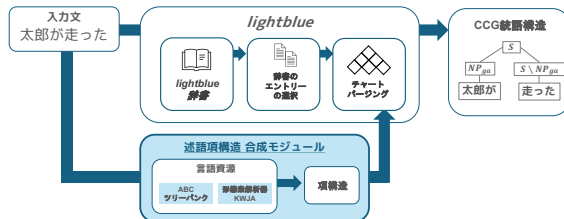


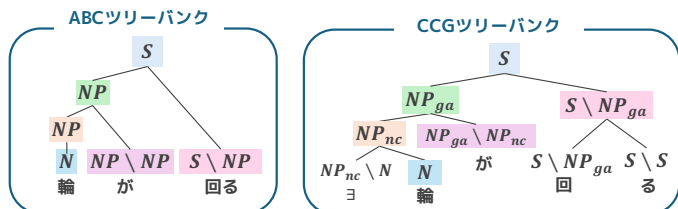
図2：項構造合成モジュールを加えたツリーバンクの構築手順

提案手法1：他コーパスとの比較

- 他の範疇文法コーパスとの一致度をスコア化することで、統語構造としての信頼性を評価する
 - 今回は、CCGへの変換が容易なABC文法に基づいたABCツリーバンク (Kubota et al. 2019) を用いる

スコアの付け方

ABCツリーバンクの部分構造が
CCGツリーバンクの部分構造に含まれる割合



✓ unary規則にも対応できる

✓ 述語の分析の差異にも対応できる

✗ 一方を「正解データ」とであると仮定する必要がある

✗ 統語素性の評価ができない

提案手法2：型検査での評価

型検査(type-check)：CCGツリーバンクに含まれる依存型意味論(Bekki 2017)に基づく意味表示が型として整合であるかを確かめることができる

型検査が成功：型レベルでの整合性は担保できる

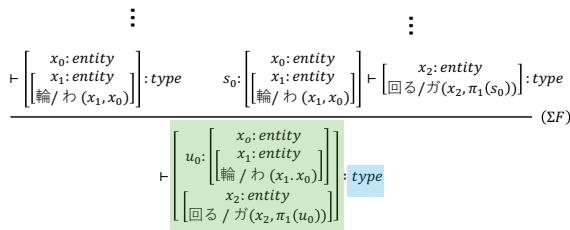


図3：「輪が回る」の意味表示の型検査の証明木の一部

意味表示がtypeという型を持つことが証明できれば成功
型検査が失敗：統語構造または意味表示に誤りがある

✓ 意味表示のレベルで型理論に基づいた評価ができる

✗ 型検査に成功した統語構造が全て妥当であるとは限らない

まとめ

- ・ CCGツリーバンクの妥当性を自動で評価するための手法を2つ提案した
- ・ 統語的な評価と意味的な評価を組み合わせることで、多角的にコーパスとしての質を担保することができる
- ・ 自動評価の結果は、実装形式文法自体の改良にも役立つ
→ パーザーや文法開発の研究とコーパス研究が相補的な関係を持つようになる