

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：コーパスの評価

- コーパス自体がどれだけ妥当であるか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

- 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



LLMでの推論ではダメなのか？

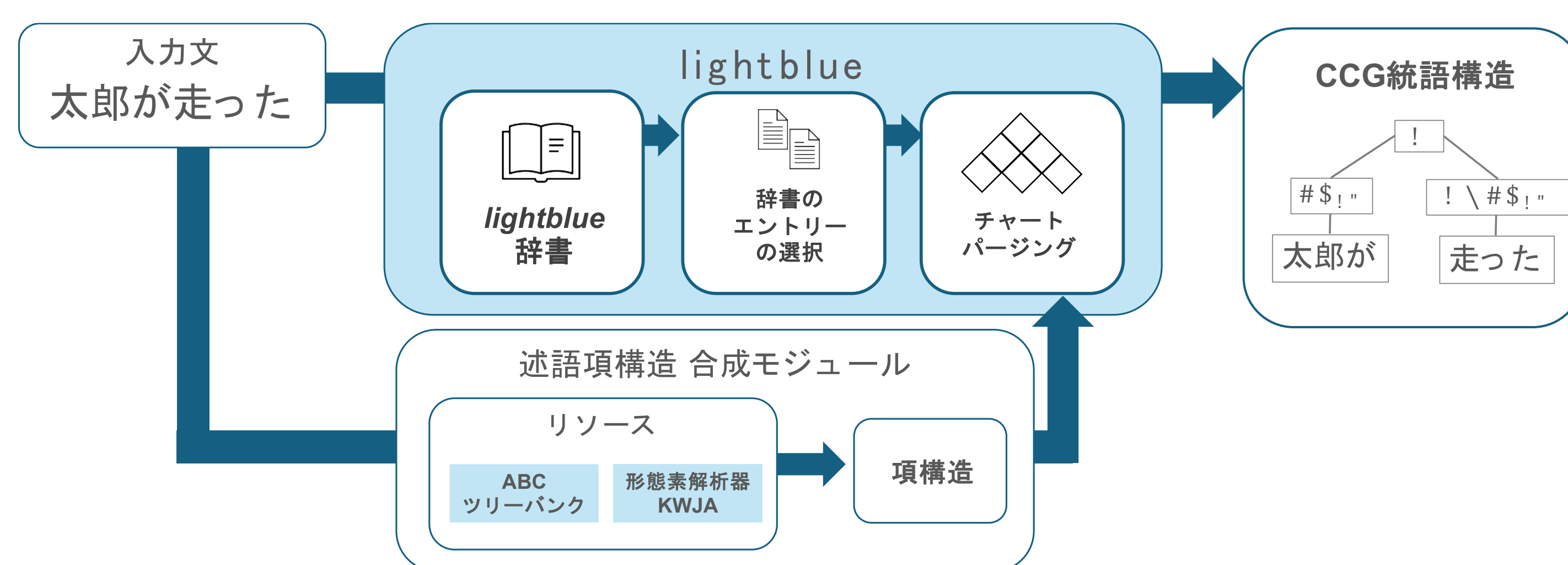
LLM：なぜ誤りが生じたのかの原因の特定が難しい

形式文法：誤りの原因の特定が可能

→ 信頼性につながる

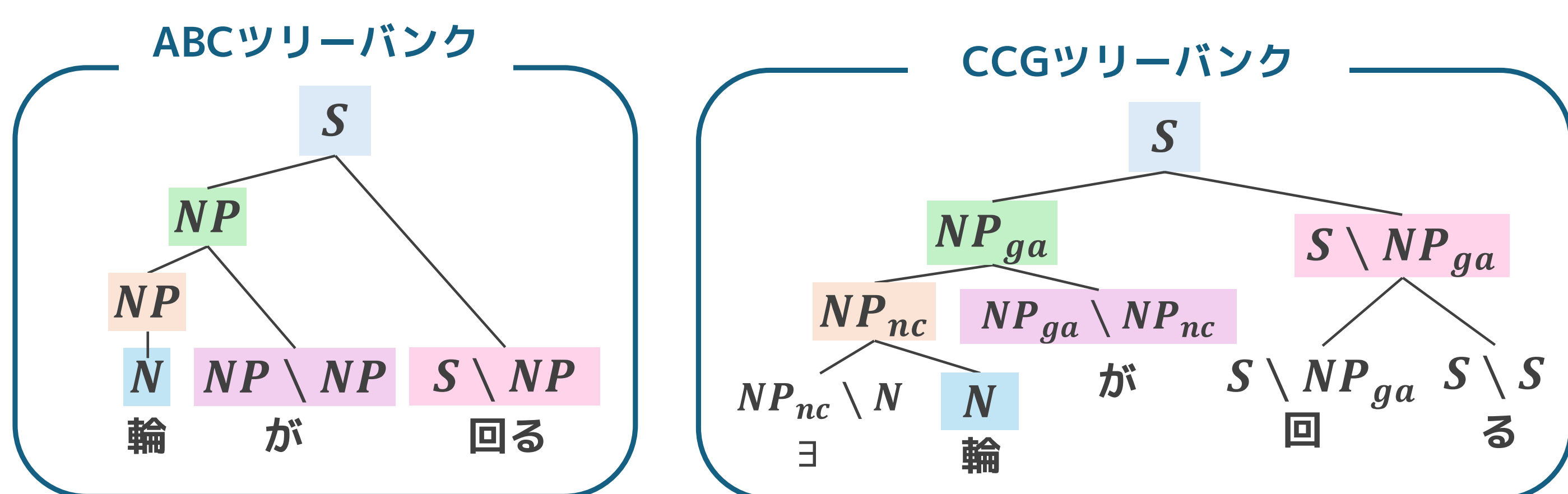
ツリーバンクの構築 (Tomita et al., EACL 2024)

- 自動でアノテーションを行うために
日本語CCG統語解析器lightblueを用いる
- lightblueが参照する辞書には、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**で他の言語資源から項構造情報を抽出し、合成する



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測ることで、どれくらい統語構造として信頼できるかを評価する
- 今回は比較対象の範疇文法として、CCGへ変換しやすい**ABC文法のABCツリーバンク**を用いる
- 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ている割合で評価



✓ 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

$$\vdash \left[\begin{array}{l} x_0: \text{entity} \\ x_1: \text{entity} \\ \text{輪} / \text{わ} (x_1, x_0) \end{array} \right] : \text{type} \quad s_0: \left[\begin{array}{l} x_0: \text{entity} \\ x_1: \text{entity} \\ \text{輪} / \text{わ} (x_1, x_0) \end{array} \right] \vdash \left[\begin{array}{l} x_2: \text{entity} \\ \text{回る} / \text{ガ} (x_2, \pi_1(s_0)) \end{array} \right] : \text{type} \quad (\Sigma F)$$
$$\vdash \left[\begin{array}{l} u_0: \left[\begin{array}{l} x_0: \text{entity} \\ x_1: \text{entity} \\ \text{輪} / \text{わ} (x_1, x_0) \end{array} \right] \\ x_2: \text{entity} \\ \text{回る} / \text{ガ} (x_2, \pi_1(u_0)) \end{array} \right] : \text{type}$$

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
- それぞれに利点・欠点があるが、評価を組み合わせることで、角的にコーパスとしての質を担保することができる