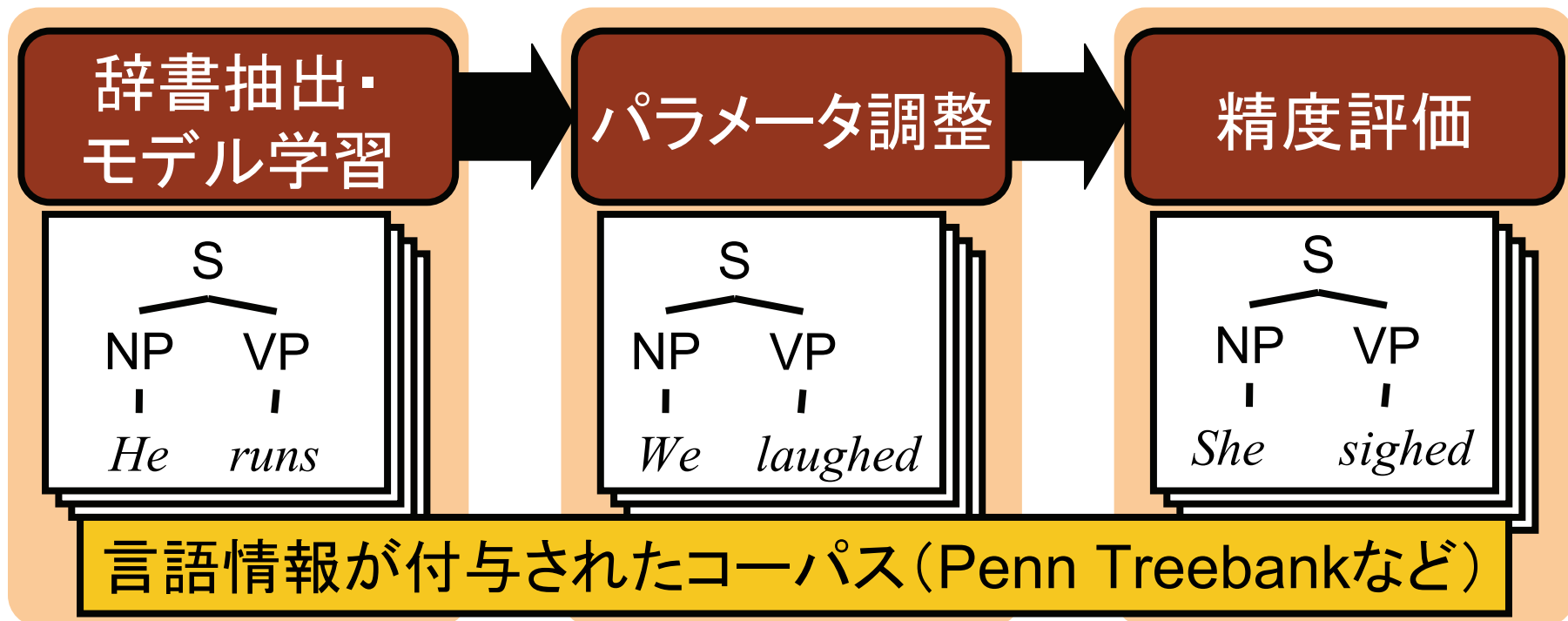


構文解析器開発実験の観点から見た コーパスアノテーションの揺れの分析

原 忠義, 辻井潤一
(東京大学)

大規模コーパスを用いた 構文解析器開発実験

- 各実験過程において、言語情報が付与されたコーパスを利用
 - 大量の実データに基づく頑健な開発実験



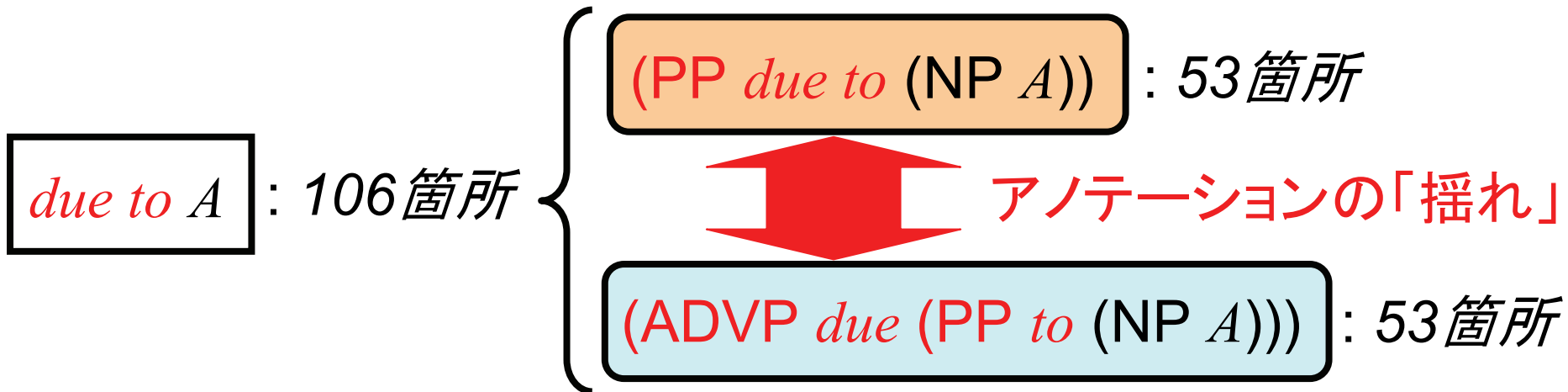
構文解析器開発実験の コーパスアノテーションへの依存性

- コーパスアノテーションの質に強く影響を受ける
 - *Enju*^{*1*2} (HPSG 構文解析器) の解析精度評価
 - 構文解析エラーの 23.2% → 正解コーパスの誤り (GENIA 50 文, 166個のエラー分析)
 - 他の実験過程にも影響を及ぼしている可能性

コーパスアノテーションの誤りが実験へ及ぼす影響を調査

コーパスアノテーションの 全体的な「揺れ」

- 各実験過程に影響を及ぼす可能性
 - 不安定なモデル学習, 不適切なパラメータ調整
 - 精度評価の不当な上昇・下降



ポスター発表

- 実際に数種の「揺れ」を手作業で修正
- 修正前と後のコーパスで実験結果を比較
→ 「揺れ」の影響を具体例で分析

「構文解析器開発実験におけるコーパスのより適切な扱い方」について考えるきっかけにしたい

