

# Grammar Error Detection with Best Approximated Parse

林 秀治

# 目的

- 最低限の誤りをなくすために、文法誤り検出時に発生する文の繋がりの曖昧性を解消する。
- 文法の適格性だけに依存し、ほかの文法資源や規則緩和、エラー処理も必要としないモデル理論的な構文のアルゴリズムを提案する

- フレーズだけで見れば完全に検出できるかもしれないが、特定の前後関係のときには誤りがあるかもしれない
- 最小のコストの強い制約でエラーを検出する
- 非文法的な入力文があれば、そこから近い木を解析し、文法制約の違反を探す

# Property Grammar

- 知識表現のために使っているフレームワークは Property Grammar(Blache, 2001)(PG)
- PG文法は統語的特性の生成文法の書き換え規則を分解する

# 解析アルゴリズム

- モデルのクラスはペアワイズ交差なしで木構造にさらに制限される
- 解析の際、1つの根を持つ必要があるので、カテゴリが見つからない場合スターというワイルドカードがノードの代わりに使われる
- スターのカテゴリは文法においてどんな制約にも指定されない

# LSCP

- Loose Satisfaction Chart Parsing(LSCP)というアルゴリズムを導入する
- LSCPはN分木を生成する
- 影響を与えるノードの子だけを見るので問題を簡単化できる

# LSCPのアルゴリズム

- 基礎準備: 語ごとに番号やカテゴリーなどのタグをつける

STEP1: 今参照するスパンを選ぶ

STEP2: 全ての構成を列挙する

STEP3: 文法からの制約システムの特徴付け

STEP4: カテゴリーへの投射

STEP5: 最適な候補を保存

# 評価

- EASY evaluation campaign of parsers of French ( Paroubek et al. , 2003 )のコーパスの一部を評価に使用し、その評価基準に従って評価した
- 比較のためにVanRullen (2005) が提案した文法でも評価した



# 評価

- 非文の誤りを見つける能力を解析するため、理想的には整っていない発話のコーパスがよいが用意できなかった
- EASY evaluation campaign of parsers of French ( Paroubek et al. , 2003 )のコーパスの一部を評価に使用し、その評価基準に従って評価した
- 比較のためにVanRullen (2005) が提案した文法でも評価した

# 結果

	Precision	Recall	F
Total	0.7835	0.7057	0.7416
general lemonde	0.8175	0.7515	0.7837
general mlcc	0.7175	0.6366	0.6746
general senat	0.8647	0.7069	0.7779
litteraire	0.8124	0.7651	0.788
mail	0.7193	0.6951	0.707
medical	0.8573	0.678	0.757
oral_delic	0.6817	0.621	0.649
questions amaryllis	0.8081	0.7432	0.7743
questions trec	0.8208	0.7069	0.7596

# 結果

	Precision	Recall	F
shallow parser	0.7846	0.8376	0.8102
stochastic parser	0.9013	0.8978	0.8995

# まとめ

- 構文木を使って影響のある子だけを選択して文法誤り検出を行うLCSPのアルゴリズムを導入した
- 今回は最適なコーパスではなかったもので、より最適なコーパスでの結果の確認と最適化が今後の課題

# 紹介論文

## **Grammar Error Detection with Best Approximated Parse**

Jean-Philippe Prost

LIFO, Universit e d'Orl ans

INRIA Lille - Nord Europe

Proceedings of the 11th International Conference on Parsing Technologies (IWPT), pages 172.175, 2009.