

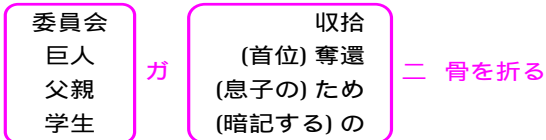
慣用句格フレームの自動構築に向けて

橋本 力
山形大学大学院理工学研究科
ch@yz.yamagata-u.ac.jp

河原 大輔
情報通信研究機構
dk@nict.go.jp

1. 慣用句格フレーム

● 慣用句用例から格フレーム (CF) を作る

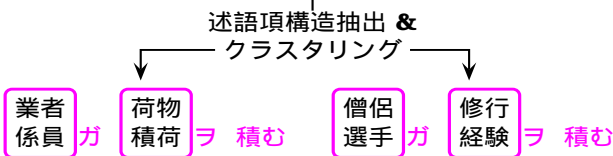


- 慣用句の用例のみを正確に収集するのが鍵
慣用句：委員会が事態の收拾に骨を折った。
リテラル句：選手が足の骨を折った。
- 佐藤 (2007) に収録の述語慣用句を対象とする
- 用例収集元は 5 億文の Web コーパス
- 慣用句検出タスクを通して評価する
- 慣用句の CF 構築はほとんど研究されてこなかった

2. 要素技術

2.1. 格フレーム自動構築 (Kawahara & Kurohashi, 2006)

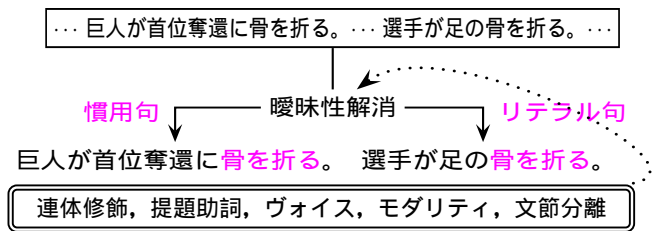
… 業者が荷物を積む。… 僧侶が修行を積む。
選手が経験を積む。… 係員が積荷を積む。…



- 本研究 → 事前に用例を慣用句とリテラルに分けるべき
 - クラスタ数は常に 2 つ (慣用句とリテラル句)
 - 同じ格スロットに入る単語の意味がより広範囲

2.2. 慣用句検出 (Hashimoto et al., 2006)

述語慣用句 { ㊦ → 曖昧性解消不要「役に立つ」
 ㊧ → 曖昧性解消必要「骨を折る」



- 文法的制約違反が無ければ慣用句と見なしてしまう
→ 曖昧性解消の精度向上には CF が必要

3. 提案手法

- ㊧ は CF 構築前に用例を慣用句とリテラルに分類。
限られた手掛かりから始め、徐々に信頼できる手掛かりを集めて曖昧性解消の精度を上げる **Bootstrap** の手法

Phase 1 文法的制約のみ

Phase 2 文法的制約とリテラル句 CF

Phase 3 文法的制約、リテラル句 CF、慣用句 CF

- ㊦ は用例分類せずに CF 構築

4. 今後の課題

- 予備実験では曖昧性解消時の失敗が目立った
→ Web 文に対する形態素 / 構文解析のロバスト化
→ Web 文に対する文法的制約の最適化

