

複合動詞・構成動詞間の意味的な共通性を計る指標の評価

山口昌也

(国立国語研究所)

概要

● 本研究の背景：複合動詞の構造分析

■ 格支配による分類(山本1984)

- V+V: 投げ捨てる
- V+v: 見上げる
- v+V: 打ち捨てる
- v+v: 繰り返す

■ 語彙概念構造(由本2005)

切る 倒す
LCS1 LCS2
⇒ LCS2 BY LCS1

意味の共通性の判断が研究者の内省に依存

構成動詞の意味的な変化が仮定されていない

● 本研究の目的

- 意味的な共通性を指標を実例に基づき評価
- 構成動詞の意味的な変化の有無を確認

評価実験

● 方法

格要素の共起ベクトルによる指標



人手による意味的關係の判定

重複度(頻度考慮あり, なし)
Simpson, Jaccard, COS

- 当該の格要素が構成動詞の格要素としても適格か?
- 当該の構成動詞・格要素の意味が複合動詞の意味の一部として適切か

● 使用するデータ

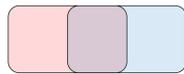
- 「複合動詞用例データベース」(複合動詞3362語, 構成動詞1040語)
- ランダムに512語抽出
用例数1000以上, かつ, フ格つきの格要素50例以上
(該当する動詞1817語)

■ 人手判定結果

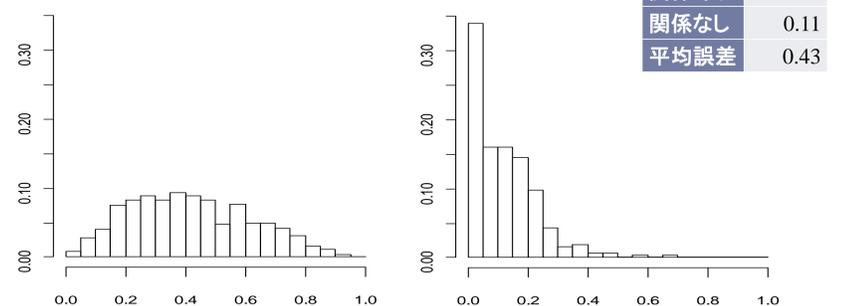
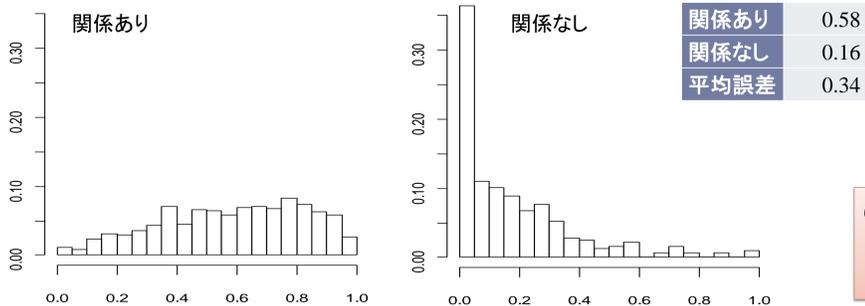
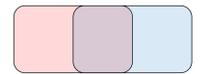
	関係あり	関係なし
前項動詞	426	86
後項動詞	269	243

結果

$$Overlap_v(w_c, w_s) = \frac{\sum_{w_a \in W_c \cap W_s} F_c(w_a)}{\sum_{w_b \in W_c} F_c(w_b)}$$

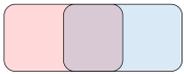


$$Overlap_s(w_c, w_s) = \frac{|W_c \cap W_s|}{|W_c|}$$

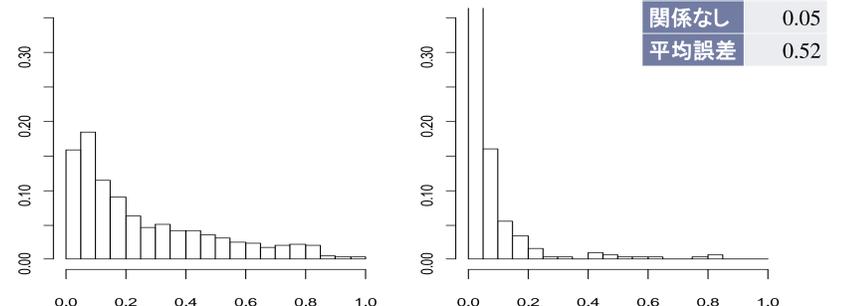


Overlap_vが最も誤差が少ない

$$Jaccard(w_1, w_2) = \frac{|W_1 \cap W_2|}{|W_1 \cup W_2|}$$



$$COS(w_1, w_2) = \frac{w_1 \cdot w_2}{|w_1| |w_2|}$$



関係あり	0.09
関係なし	0.02
平均誤差	0.63

考察

● 構成動詞が誤差に与える影響

構成動詞にしか出現しない格要素 ⇒ 誤差

■ 構成動詞の多義性

「買い集める」と「買う」
 ・購入する ○車を買集める
 ・評価する ×努力を買集める
 ・引き受ける ×けんかを買集める

■ 複合による格要素の制限

「飲み明かす」と「飲む」 ⇒ 一般的に酒類に限定
 ○酒を飲み明かす
 ×ジュースを飲み明かす

● 複合時の意味的变化

■ 語義の派生

「打ち破る」の例
 ・壁を打ち破る ⇒ ○壁を打つ
 ・現実を打ち破る ⇒ ×現実を打つ

■ 格要素の出現確率の変化

構成動詞側では、あまり用いない格要素が複合動詞側で使用される

格要素	焼き払う	焼く	差
村	0.055	0.0	0.055
敵	0.047	0.012	0.035
町	0.032	0.0	0.032
家屋	0.025	0.0	0.025
寺	0.041	0.16	0.025

■ 人手判定との差異

- 前項・関係あり, $Overlap_v \leq 0.3$ (47)
 - 複合動詞の多義性(57.4%)
 - 出現確率の変化(53.2%)
- 前項・関係なし, $Overlap_v > 0.3$ (18)
 - 構成動詞の多義性(61.1%)
 - 誤用(22.2%)
- 後項・関係あり, $Overlap_v \leq 0.3$ (51)
 - 出現確率の変化(74.5%)
 - 複合動詞の多義性(27.5%)
- 後項・関係なし, $Overlap_v > 0.3$ (46)
 - 構成動詞の多義性(80.4%)
 - 解析誤り(13.0%)