

日本語のガヲニデなどの格の分類 による文からの表現の構造化

岡山大学

大学院自然科学研究科

産業創成工学専攻

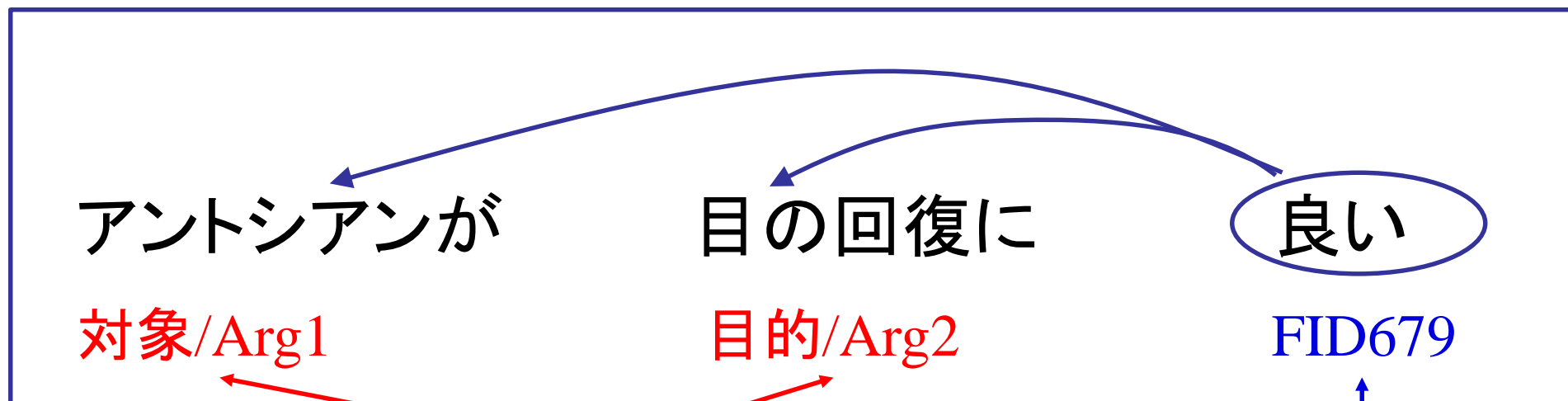
講師

竹内 孔一

令和2年9月17日

格を意味関係に分類した解析

述語の概念を識別し意味関係を同定する



意味役割ラベル (格関係の意味的なタイプ分け)

概念フレーム (可能)

他の表現

アントシアンが	目の回復に	効く/効果的だ
対象/Arg1	目的/Arg2	FID679

同様の表現
との対応関係
が取れる

従来の格解析の問題点

同様の文節と述語の意味関係を同定する先行研究にKNP、ChaPasがあるが、

- 「が」「を」「に」の3つの格のみ対応(ChaPas)
- 「岡山を出発する/出る」など言い換え関係が不明
- 動作主やArg1など意味役割ラベルがない

など、述語と係り元の関係を簡単なラベルで捉えることができていない。

新技術の特徴・従来技術との比較

- 従来技術の問題点であった、意味分割と概念フレームを規格化した辞書を無償公開。
- 従来の辞書データは基本有料であったものを、Web上で無償公開することで、WordNetと同様に分類を誰でも確認することができる。
- 自動解析システムの適用により、異なる表現を辞書の観点から整理して抽出できる。

意味役割と概念フレーム

仮説: 各概念には固定の登場要素がある



[動作主/Arg0 住民が] [基準/Arg2 建設に] [FID166 反対する]

[動作主/Arg0 住民が] [基準/Arg2 建設に] [連語 異議が] [FID166 ある]

「～がある」など構文の異なりも構造化

これで1つの述語

構築している辞書と解析システム

述語項構造シソーラス

- 各述語の例文データ (<http://pth.cl.cs.okayama-u.ac.jp/>)
- 1.1万語の述語、約1100の概念、約2300例文、約80種の意味役割

例)「走る」は「線路を走る」「悪事に走る」「ストレートが走る」
など語義の違いを記述。「口が滑る」など慣用句も登録。

意味役割解析システム(ASA)

- シソーラス辞書の事例を本に文の格を解析して
- pythonの実装を公開

(<http://www.cl.cs.okayama-u.ac.jp/study/project/asa/>)

述語項構造シソーラス

Id ↓	見出し語	Yomi	大分類1	大分類2	中分類	小分類1
5440	現わす	アラワス	状態変化なし (状態)	主体の性質	社会的状態 (人)	社会的位置づけ
	その企業が [経験者][Arg1] 名を [対象 (身体部分)] [Arg2] 世間に [領域 (人)] [Arg3] 現わす					
5451	ときめく	トキメク	状態変化なし (状態)	主体の性質	社会的状態 (人)	社会的位置づけ
	新人作家が [経験者][Arg1] 今を [時間][ArgM-TMP] ときめく					
13554	知られる	シラレル	状態変化なし (状態)	主体の性質	社会的状態 (人)	社会的位置づけ
	彼は [経験者][Arg1] 俊足で [補語相当 (が)] [Arg2] 知られる					
13993	あっぱれだ	アッパレダ	状態変化なし (状態)	主体の性質	社会的状態 (人)	社会的位置づけ
	結果が [対象 (身体部分)] [Arg2] あっぱれだ					

Koichi Takeuchi, Alastair Butler, Iku Nagasaki, Takuya Okamura, Prashant Pardeshi, Constructing Web-Accessible Semantic Role Labels and Frames for Japanese as Additions to the NPCMJ Parsed Corpus, In Proceedings of 12th Edition of its Language Resources and Evaluation Conference (LREC2020), 2020.

意味役割解析システムASA

今すぐASAを試しましょう。

恵まれていたので、彼は図に乗っている

Submit

Graph Raw

Idiom

図に

恵まれていたので、

彼は

乗っている

[List(着点)]

[List(原因)]

[List(対象)]

Koichi Takeuchi, Suguru Tsuchiyama, Masato Moriya, Yuuki Moriyasu and Koichi Satoh,
Verb sense disambiguation based on thesaurus of predicate-argument structure,
In Proceedings of International Conference on Knowledge Engineering and Ontology Development,
pages 208-213, 2011

想定される用途

- 質問応答やチャットボットにおける類似表現の検索に適用することで、用意された辞書ベースの言い換え表現とマッチして、類似文書を抽出できる。
- 文の表現を「**動作主**」「**対象**」「**時間**」など一定の関係に構造化できるため、文から「誰が」「何を」「いつ」したかなど一貫した関係を取り出せる。

実用化に向けた課題

- 現在、1.1万語の述語(動詞・形容詞・形容動詞)について60%の精度で意味役割付与できる部分まで開発済み。しかし、まだ精度が低い点が未解決である。
- 現在、国立国語研究所と6万文の基本データを共同で作成中。データを利用して精度向上を目指している。
- 実用化に向けて、精度を90%まで向上できるように技術を確立する必要がある。

企業への期待

- 意味役割付与の解析精度の向上は、現在構築中の基本データで改善することが期待できる。
- 電子化された文書を蓄積し、文書に対する質問応答システムの精度の改善を求める企業との共同研究を希望。
- 近年の深層学習を利用したベクトルベースの自然言語処理モデルで解析に問題を抱えている場合、本技術の導入が有効と思われる。

他機関と連携実績

- 2007年 - 2008年 共同研究の実施(NTT)
- 2019年 共同研究の実施(企業)
- 2019年 - 2020年 受託研究の実施
(国立国語研究所)

お問い合わせ先

**岡山大学 研究推進機構
産学連携・知的財産本部**

TEL: 086-251-8463

FAX: 086-251-8961

e-mail: cr-ip@okayama-u.ac.jp