



OKAYAMA
UNIVERSITY

日本語の述語を中心とした 文書の構造化技術

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

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
対象/Arg1

目の回復に
目的/Arg2

効く/効果的だ
FID679

同様の表現
との対応関係
が取れる

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。

- 従来技術の問題点であった、**意味役割と概念フレーム**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。

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]

これで1つの述語

「～がある」など構文の異なりも構造化

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/>)

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 (か)][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					

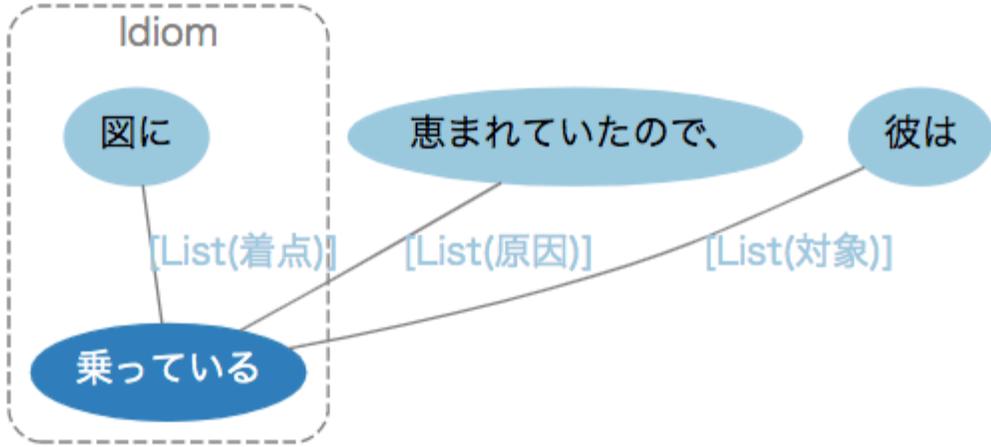
Koichi Takeuchi, Alastair Butler, Iku Nagasaki, Takuya Okamura, Prashant Pardeshi, Constructing Web-Accessible Semantic Role Labels and Frames for Japanese as Additions to the NPCMJ Parsed Corpus, In Proceedings of 12th Edition of its Language Resources and Evaluation Conference (LREC2020), 2020.

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



```
graph TD; A(恵まれていたので、) ---|"[List(原因)]"| B(乗っている); C(彼は) ---|"[List(対象)]"| B; D(図に) ---|"[List(着点)]"| B; subgraph Idiom; B; end
```

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

- 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。

述語の意味を構造化して検索

例) 意見文書からの分析

質問

ユーザが何に反対しているか?

述語の構造化によるクエリ

反対(ユーザ, ?Arg1, ?Arg2)
→ Arg2 を文書から取り出す



検索結果

.. [Arg2 値上げに] 異議があったり..
.. [Arg2 オプションの廃止に] 強く反対..

意味役割解析システムASAによる構造化

述語が持つ概念の登場要素で意味を整理

述語シソーラス形式による構造化



→「強く」など挿入があっても簡単に表現が抽出検索できる

- 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な技術を確立する必要がある。

- 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。

- 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)



**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



ご清聴ありがとうございました