

日本語テキストに対する述語語義と 意味役割のアノテーション

言語処理学会第18回年次大会(NLP2012)

@広島市立大学

3月14日(水) 15:00-18:00 502講義室

林部 祐太⁺ 小町 守⁺ 松本 裕治⁺

⁺奈良先端大

隅田 飛鳥^{*}

^{*}KDDI 研究所

述語項構造解析

項

述語



私は

妻に

花を

贈った。

ガ格

二格

ヲ格

関係ラベル

機械翻訳 [Wu 09] や情報抽出 [Surdeanu 03]などに有用

項と述語の関係の粒度 (日本語)

表層格レベル(助詞相当ラベル)

ガ格 二格 ヲ格

従来の多くの研究の
対象

[河原 07, 飯田10, Taira 10, 吉川 10]

深層格レベル(意味役割ラベル)

対象 動作主 経験者

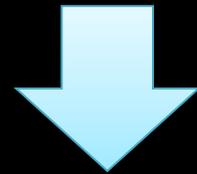
[肥塚 07, 竹内 10]

あまり対象にされてこ
なかった

理由の1つ:
実テキストへ意味役割を大規模にアノテーションしているコーパスが未整備
(英:PennTreebank, 英中亜:Ontonote)

本研究の目的

意味役割



アノテート

日本語テキスト

意味役割付与の応用例

・言い換え

同一事象を表すが、統語的に異なる構造持つ場合
→深層格なら同じ項構造

風呂が水で一杯になった

place

object

水が風呂を一杯にした

object

place

[辻井 85]

意味役割



依存関係

語義

[渡邊10]

語義決定に意味役割情報が必要な場合

She drives a car.

drive.01 : object=vehicle

drive.02 : object=things in
motion

[渡邊10]

PropBankによる

意味役割決定に語義情報が必要な場合

She comes to work on weekends.

[A1]

come.01 : 動作を意味

She comes to terms with it.

[A0]

come.14 : 妥協する

[渡邊10]

PropBankによる

本研究の貢献

- **述語の項に意味役割を付与**

- 表層格レベルで既に付与されているNAIST テキストコーパス [飯田10]に対して
- 動詞項構造シソーラス[竹内08, 竹内11] で定義されている項構造と意味役割の体系に則った

- **述語の語義ラベルも付与**

- 述語の語義と意味役割には相互に依存関係がある[渡邊10]ことから

語義・意味役割付与済み日本語コーパス

コーパス名	対象データ	意味役割の付与対象	語義辞書
動詞項構造シソーラス	作例 7,473 文	文内項	学研 Lexeed
EDR 電子化辞書	新聞, 雑誌, 辞典等の約 20 万文	文内項	EDR 概念辞書
GDA[橋田 05] コーパス	1994 年毎日新聞約 3.7 万文	文間項, 外界項	岩波国語辞典第五版
[小原 11] のコーパス	BCCWJ(現代日本語書き言葉均衡コーパス)840 文	全ての自立語にフレーム名	-
今回作成したコーパス	1994 年毎日新聞 2338 事例	<u>文内項, 文間項, 外界項</u>	学研 Lexeed

深層格レベルでの項構造の体系と 辞書

深層格レベルでの項構造の体系と辞書

フレーム意味論に基づく項構造体系

概念意味論に基づく項構造体系

概念意味論に基づく項構造体系

- 語彙概念構造(Lexical Conceptual Structure; LCS)[Jackendoff 90, 影山96]
 - “BECOME”, “CONTROL” のような意味述語を意味特性ごとに用意
 - それらの組み合わせで動詞の意味を記述

表 1: T-LCS 辞書中の LCS の例

ID	LCS LCS のインスタンス	動詞の例 例文
10	[y BE AT z] [本 BE AT 棚]	位置する, 存在する 本が棚に存在する
7	[y MOVE TO z] [本 BE AT 棚]	移動する, 遷移する 本が棚に移動する
2	[x CONTROL [BECOME [y BE AT z]] [朋子 CONTROL [BECOME [本 BE AT 棚]]]	設置する, 置く 朋子が本を棚に置く

動詞項構造シソーラス[竹内08, 竹内11]

- [竹内06] によるLCS辞書を用いた
- 動詞を横断した最小限の意味役割を87 種類設計
- 動詞4,425語(7,473 語義)の作例に対して付与

本研究のアノテーション方針

NAISTテキストコーパス1.4β

- 日本語述語項構造解析のタスクの訓練と評価において広く用いられているコーパス
- 京都大学テキストコーパスVersion 3.0を元に
- 約4 万文に対して、述語の格関係、事態性名詞の格関係、名詞間の照応関係がアノテーションされている

付与対象

- 対象の動詞
 - 中頻度の動詞
- 対象の項
 - 各動詞の必須格の項に絞った
 - 全ての格を対象にすると付与すべき項の数が増え、付与できる事例数が減少するため
- 項は種類(文内項, 文間項, 外界項)に関係なく全て対象に付与

付与したラベル

- 動詞項構造シソーラス(LCS)[竹内08]とそこで定義されている意味役割を付与
 - 既存の辞書の中で最も整備が進んでいる
 - 項構造の検索システムも整備されている
- 「副詞相当」等の文法機能は除外
- 「結果物」と「生成物」のように区別のつきにくい意味役割は1つにまとめた
- 大量の意味役割の中から1つを選ぶのは困難
 - 意味役割を役割ごとにグループ化し, グループを選択した後に, 意味役割を付与するようにした.

表 3: 付与した意味役割とその数

グループ	意味役割
対象 (1,947)	対象 (1279), 対象:事態 (406), 対象:人 (145), 対象?(74), 対象:身体部分 (30), 対象:生物 (13)
動作主 (1,303)	動作主 (1,226), 動作主?(72), 動作主:操作対象 (5)
起点・着点 (299)	着点 (178), 着点?(45), 着点:場所 (35), 着点:身体部分 (20), 着点:人 (20), 起点:場所 (1), 起点・着点 (0)
経験者 (268)	経験者 (177), 経験者?(91)
モノ (94)	生成物 (82), 内容物 (5), 材料 (5), 内容物?(2)
点・方向 (64)	基準点 (38), 方向 (12), 方向?(7), 経由点 (4), 経路 (2), 方向:人 (1)
はたらき (52)	役割 (26), 状態 (15), 状況 (9), 状況?状態?(2)
原因・結果 (26)	原因 (23), 結果 (1), 決定内容 (2)
場所 (24)	場所 (20), 範囲 (4), 境界 (0)
程度 (21)	程度 (21), 数量 (0)
目的 (14)	目標 (7), 用途 (6), 目的 (1)
時 (11)	期限 (6), 時 (3), 期間 (2)
手段 (6)	道具 (6), 手段 (0)
変化 (3)	変化前 (1), 変化先 (1), 変化先?(1)
その他 (171)	慣用 (59), 感情 (30), 削除 (48), その他 (34)

語義ラベルの付与

- 語義と意味役割には相互に依存関係がある[渡邊10]
 - 意味役割付与と語義曖昧性解消の同時学習の研究を可能とするため
- 「基本語データベース:Lexeed」[笠原04] の語義番号を付与
 - 動詞項構造シソーラスでも用いられている
- 慣用句の一部として用いられている場合
 - 慣用句の特殊ラベルを付与
- Lexeedには無い語義で用いられている場合
 - 「新語義」の特殊ラベルを付与

作成したデータの分析

統計

- 計172種類, 延べ2,338個の動詞を対象
 - 語義とその項の意味役割を付与
- 総意味役割付与数は4,398個
 - 上位3グループの付与数は全体の80%以上を占める
 - この偏りは, 付与対象を必須格に限定したことが理由
 - カラ格を付与対象にすれば, 「場所」や「方向」
 - デ格を付与対象にすれば, 「道具」や「手段」の意味役割は増える

ガ格の意味役割の分布

- 大半は次の3つが占めている
 - 操作対象に対して直接ある操作を行う「動作主」
 - 意志性が無いものが主体となる「対象」
 - 意志性がある者が不本意な状況に陥る主体となる「経験者」
- 「対象:事態」は「コメ生産」や「新党派結成」といった事態性名詞が項となるときに付与

ガ格 (2,277)	動作主(1,226), 対象(450), 経験者(177), 対象:事態(125), 経験者?(91), 動作主?(72) ...
上位6種のカバー率	94.0%

ヲ格の意味役割の分布

- 「対象」(とその細分類)が多い
 - 行為の直接対象を表す意味役割
- 「生成物」
 - 「その行為の結果生じるもの」で「対象」とは区別
 - ヲ格項を持つ124 動詞中, 「施行する」・「組織する」・「作り出す」などの8 種類の「何かを生成することを表す動詞」で用いられた
- 慣用句の一部として用いられていることを示す意味役割「慣用」も他の格と比べて多い

ヲ格 (1,421)	対象(761), 対象:事態(220), 対象:人(108), 生成物(82), 慣 用(38), 削除(35) ...
上位6種のカバー率	87.5%

二格の意味役割の分布

- 意味役割の偏りが最も小さかった
- また、「方向」「期限」、「内容物」など二格にしか現れなかった意味役割もいくつかあった

二格 (6,05)	着点(178), 対象(68), 対象:事態(61), 着点?(43), 着点:場所(35), 対象:人(29)...
上位6種のカバー率	68.4%

語義ラベルの一致率

- 8 動詞92 事例を2 人の作業者でアノテーションした
- 語義ラベルの一致率は68%
 - 29 事例一致しなかった
 - 慣用句か否かの見解の違い 8事例
 - 新語義か否かの見解の違い 6事例
 - 語義の認識の違い 14事例

意味役割の一致率

- 一方の作業者のアノテーションを正解
- もう一方をシステムの出力とみなす
- 意味役割の一致率はガ格, ヲ格, 二格の順に低下する
- 意味役割の分布のばらつきもこの順に大きくなる
- ラベルの選択候補が増えることが理由の1つ

	P	R	F
ガ	0.98 (59/60)	0.65 (59/91)	0.78
ヲ	0.83 (24/29)	0.41 (24/58)	0.55
二	0.36 (4/11)	0.25 (4/16)	0.30

アノテーションの問題点と対応策

比喩的表現

- 語義ラベルの決定
 - 動詞と項だけでなく文脈を考慮する必要がある
- 「風に吹かれるままに」
 - 「吹く」に最も近いLexeedの語義は「風が物を揺り動かしながら通っていく」
- 「運命の風に吹かれるままに」
 - ある作業者は「吹く」に「新語義」ラベルを付与

比喩的表現かどうかの判断の揺れ

- 比喩的に用いられているかの判断
 - 作業者間での揺れ
- 比喩的な表現の可能性がある場合
 - 最も近い語義を選択
 - さらに「比喩的表現」ラベルを付与する

意思性の有無

- 主体に対する意味役割
 - その主体の意思性の有無によって「動作主」か「経験者」が変わる
 - 判断には揺れが生じる
 - 「？」が付いたラベルは「経験者」が最も多かった

意思性の有無の判断の揺れ

- 「香港は... 経済発展を遂げた。」
 - 「香港」が、意思をもって経済発展を「遂げた」？
 - 偶然経済発展を「遂げた」？
 - この文からは判別できない
- 「経験者？」や「動作主？」といった意味役割レベル
 - 曖昧さを表現するために「経験者」や「動作主」とは独立して用いる

まとめ

- NAIST テキストコーパスに対して語義と意味役割のアノテーションを行った
 - 体系的な大規模アノテーションを行うことで、意味役割の頻度分布が分かった
 - 用いられやすい意味役割や、格による意味役割の使用頻度の偏りが分かった
- 今後の課題
 - さらなるデータの分析を行い、アノテーション基準を洗練させる
 - 必須格以外への意味役割のアノテーション

APPENDIX

タグ付システム

- 専用のウェブサイトを構築

語義タグ付システム

お知らせ

- **new!** 意味役割タグ付け完了データダウンロードできるようになりました(2011.6.14)

システムメニュー

- [パスワードの変更](#)
- [DBの管理](#)
- [SHA512を求める](#)
- [SHA512を一括で求める](#)
- [動詞の頻度一覧\(頻度11以上のもの\)](#)
- [意味役割タグ付け完了データダウンロード](#) [意味役割csv\(最新版\)](#) [語義csv\(最新版\)](#) [Excel\(csvを2011/6/14に変換したものです\)](#)

メニュー

- [タグ付対象動詞の一覧\(頻度11以上でLexeedに語義が登録されているのもの\)](#)
- [自分のタグ付け完了データ](#)
- [全てのメンバーのタグ付け履歴](#)
- [アノテーション済みデータから語義IDで検索](#)

- [岡大LOSIにおける意味役割](#)
- [岡大LOSIにおける意味役割\(最新版\)](#)
- [修正版1](#)
- [修正版2](#)

検索

- [NAISTテキストコーパスからの動詞検索](#)

- [Lexeedからの曖昧検索](#)

表記:

- [動詞項構造シソーラス\(岡山大\)](#) (登録されている動詞5190種類)

タグ付対象動詞一覧

動詞	NAISTテキストコーパス内での頻度	Lexeedでの語義数	あなたの付与件数	全てのメンバーの付与件数
突く	11	14	11	11
降りる	11	13	12	12
吹く	11	12		14
巻く	11	10		13
継ぐ	11	10		10
明ける	11	9		12
寝る	11	7		11
拾う	11	6		11
立ち上がる	11	6	11	11
浮く	11	6		11
逃す	11	6		11
乗り込む	11	4		11
脅かす	11	4		13
絶える	11	4		11
浴びせる	11	3	10	11
引き受ける	11	3		11
祝う	11	3		12
手伝う	11	2		11
困む	11	2		11
見越す	11	2		11
貸す	11	2		11
連ねる	11	2		11
劣る	11	1		11
カバーする	11	0	10	10
吸収する	11	0		11
閉鎖する	11	0		11
合流する	11	0		11

語義ラベルが不一致だった例

- 29 事例一致しなかった
 - 慣用句か否かの見解の違い 8事例
 - 「隙を突く」
 - 新語義か否かの見解の違い 6事例
 - 「雨が降る」
 - 語義の認識の違い 14事例

意味役割のグループごとの一致率

- 意味役割のグループごとの一致率
 - 意味役割ごとの一致率より上昇

	P	R	F
ガ	0.98 (65/66)	0.71 (65/91)	0.83
ヲ	0.89 (41/46)	0.71 (41/58)	0.79
ニ	0.46 (6/13)	0.38 (6/16)	0.41

ガ格の意味役割のリスト

- 動作主 1226
- 対象 450
- 経験者 177
- 対象(事態) 125
- 経験者? 91
- 動作主? 72
- その他 31
- 対象? 28
- 感情 15
- 原因 12
- 対象(人) 9
- 対象(身体部分) 6
- 場所 6
- 動作主(操作対象) 5
- 慣用 5
- 対象(生物) 4
- 状態 3
- 決定内容 2
- 内容物? 2
- 範囲 2
- 道具 2
- 状況?状態? 1
- 期間 1
- 程度 1
- 状況 1

ヲ格の意味役割のリスト

- 対象 761
- 対象（事態） 220
- 対象（人） 107
- 生成物 82
- 慣用 38
- 削除 35
- 対象？ 30
- 対象（身体部分） 23
- 役割 22
- 感情 13
- 基準点 13
- 状態 12
- 対象（生物） 9
- 状況 7
- 程度 6
- 用途 6
- 目標 6
- 材料 5
- 場所 5
- 道具 4
- 経由点 3
- 時 3
- 経路 2
- 着点？ 2
- 範囲 2
- 結果 1
- 状況？状態？ 1
- 起点（場所） 1
- 期間 1
- 変化前 1

二格の意味役割のリスト

- 着点 178
- 対象 68
- 対象(事態) 61
- 着点? 43
- 着点(場所) 35
- 対象(人) 29
- 基準点 25
- 着点(身体部分) 20
- 着点(人) 20
- 慣用 16
- 対象? 16
- 程度 14
- 削除 13
- 方向 12
- 原因 11
- 場所 9
- 方向? 7
- 期限 6
- 内容物 5
- 役割 4
- その他 3
- 感情 2
- 状況 1
- 対象(身体部分) 1
- 目標 1
- 経由点 1
- 方向(人) 1
- 目的 1
- 変化先? 1
- 変化先 1

フレーム意味論に基づく項構造体系

「フレーム」

状況や語の意味を理解するための基本単位 [Fillmore 82]

(例) フレーム「商取引」

SELLER, BUYER, GOODS, MONEY といった要素(フレーム要素)が sell や pay 等の動詞の項となる

FrameNet

フレームを想起する
語を複数挙げている

Ingestion

eat.v
drink.v
...

意味役割

Ingestibles [Ingible]
Ingestor [Ing]
Place [Place]

項と述語の2語間の
関係ではなく、項とフ
レームの関係

Inherits from: Ingest_substance,
Manipulation

Food, Traveler などの
746種類(FrameNet
1.3)

フレーム間の親子
関係7種類

日本語では「日本語FrameNet」[小原 05]

PropBank

Predicate: eat

Roleset id: eat.01 , consume, vncls: 39.1, framnet: Ingestion

eat.01: Member of Vncls eat-39.1-1.

Roles:

Arg0: consumer, eater (vnrole: 39.1-agent)

Arg1: meal (vnrole: 39.1-patient)

Example: intransitive

TW's pitch was that sales and earnings at its restaurants have risen steadily and that people-1 won't stop *trace*-1 eating during a downturn.

Arg0: *trace*

Rel: eating

Example: transitive

Euro Disneyland shares made a debut like Snow White yesterday while most of the London stock market looked like it had eaten the Evil Queen's poisoned apple.

Arg0: it

Rel: eaten

Arg1: the Evil Queen 's poisoned apple

フレームは動詞項構造ごとに用意

意味役割もフレームごとに定義(フレーム間での互換性無)

フレーム間の関係は未定義

時間を表すAM-TMP や、場所を表すAM-LOC 等のフレーム間で共通する意味役割を別途14種類定義