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

林部 祐太　　小町 守　　松本 裕治　　隅田 飛鳥
{yuta-h, komachi, matsu}@is.naist.jp　　as-sumida@kddilabs.jp
奈良先端科学技術大学院大学　　株式会社 KDDI 研究所

はじめに

述語項構造解析とは，述語に対する項を同定し，それぞれの項と述語の関係（項構造）を解析するタスクであり，機械翻訳 [Wu 09] や情報抽出 [Surdeanu 03] などの自然言語処理の応用において有用性が報告されている．

日本語テキストに対する述語項構造解析の多くは，項と述語の関係の粒度を研究されて [河原 07, Taira 10, 吉川 10] おり，項の意味上の働きを表す意味役割（深層格レベル）ラベルでの研究は，[肥塚 07, 竹内 10] など数えるほどしかない．

その理由の 1 つとして，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であることが挙げられる．本研究では，日本語において意味役割付与を行うため，述語項構造が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て項の意味役割のアノテーションを行った．また，述語の語義と意味役割には相互に依存関係がある [渡邉 10] ことから，述語の語義ラベルも同時に付与した．

本稿では，アノテーションしたコーパスの分析と，アノテーションの際に判断が難しかった事例を分析する．

関連研究

2.1 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は，既に色々と提案されている．ここでは 2 つの記述体系とそれに則って作られた辞書について述べる．

フレーム意味論に基づく項構造体系

フレーム意味論では状況や語の意味を理解するための基本単位として（意味）フレームを用いる [Fillmore 82]．例えば，「商取引」というフレームは，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ことによって構成される．

FrameNet[1] [Baker 98] では各フレームを想起する語を複数挙げている．また，フレーム間の親子関係の種類を “Inheritance”, “Perspective on”, “Subframe”, “Precedes”, “Inchoative of”, “Causative of”, “Using” の 7 つ定義している．FrameNet では，意味役割は項と述語の 2 語間の関係ではなく，項とフレームの関係を表す．FrameNet 1.3 では，Food, Traveler などの 746 にものぼる意味役割名が定義されてる．日本語においても，日本語 FrameNet[2] [小原 05] として同様の語彙情報資源の構築が進められている．

PropBank[Palmer 05] ではフレームを動詞の項構造ごとに用意している．FrameNet とは異なり，フレーム間の関係は定義していない．意味役割もフレームごとに定義しており，フレーム間での互換性は無い．なお，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 種類定義している．

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

概念意味論に基づく項構造体系

概念意味論では，動詞が表す概念的意味 (conceptual meaning) を抽象的な述語概念で表示するために，語彙概念構造 (Lexical Conceptual Structure; LCS)[Jackendoff 90, 影山 96] を用いる．具体的には “BECOME”, “CONTROL” のような意味述語を意味特性ごとに用意し，それらの組み合わせで動詞の意味を記述する．LCS は階層構造をもつ．

日本語では [大石 95] の LCS 一覧表[3]や [竹内 06] の T-LCS[4]がある．T-LCS を用いて LCS の例を表 1 に示す．x,y,z は項を示すラベルである．LCS10,7,2 の間には包含関係がある．

[竹内 08, 竹内 11] は，動詞を横断した最小限の意味役割を 87 種類設計し，T-LCS を用いて，動詞 4,425 語（7,473 語義）の作例に対して付与した．これは動

[1] http://framenet.icsi.berkeley.edu/

[2] http://jfn.st.hc.keio.ac.jp/

[3] http://www.hino.meisei-u.ac.jp/is/oishi/LCS/LCS.htm

[4] http://cl.it.okayama-u.ac.jp/rsc/lcs

表 2: 語義・意味役割付与済み日本語コーパスの比較

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	文内項，文内項，外界項	学研 Lexeed

詞項構造シソーラス[5]として公開されている．

2.2 語義・意味役割付与済み日本語コーパス

語義や意味役割が付与されている日本語コーパスを表 2 にまとめた．

項は述語との相対的位置によって，文内項，文間項，外界項に分類されるが，コーパスによって，意味役割が付与されている項は異なる．既存のコーパスでは，そのうちの一部しか意味役割が付与されていない．

語義と意味役割のアノテーション

3.1 ベースとなるコーパス

NAIST テキストコーパス 1.4β[飯田 10] は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である．これは，京都大学テキストコーパス Version 3.0[6]を元に，1995 年 1 月 1 日から 17 日までの全記事（約 2 万文）と 1 月から 12 月までの社説記事（約 2 万文）の計約 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．

京大コーパスでは述語の出現形に対して，ガ/ヲ/ニ/カラ/ヘ/ト/ヨリ/マデなどの格助詞相当のラベルや，ニツイテのような連語のラベルが付与されている．一方 NAIST テキストコーパスでは，述語の基本形（格交替は原形に戻す）の必須格（述語の概念構成に必須とされる要素; ガ・ヲ・ニ格）に対して，表層格レベルでラベル付けされている．

意味役割のアノテーションを行う際には，格交替は原形に戻す必要があるため，既にその作業が終わっている NAIST テキストコーパスをベースとなるコーパスとして用いた

3.2 実際に付与したラベル

意味役割の付与対象は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少なくなってしまうと考えたからである．項はその種類（文内項，文間項，外界項）に関係なく全て対象に付与した．既存のコーパスでは外界項や文間項のアノテーションがされていなかったが，これら全てに付与することで，項の位置に関係なく意味役割付与の訓練・精度評価が可能になる．

[5]http://cl.it.okayama-u.ac.jp/rsc/data

[6]http://nlp.ist.i.kyoto-u.ac.jp/index.php?京都大学テキストコーパス

項構造と意味役割の体系には動詞項構造シソーラス (LCS) とそこで定義されている意味役割[7]を付与した．これは既存の辞書の中で最も整備が進んでおり，項構造の検索システム[8]も整備されていることが理由である．ただし，「副詞相当」等の文法機能は除外し，「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．表 3 に実際に付与したラベルの一覧を示す．ラベル名が「〜?」で終わっているラベルは，その意味役割ラベルを付与する際に，作業者が多少の疑問を感じたときに付与した．また，大量の意味役割の中から 1 つを選ぶのは困難であるため，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．

さらに，語義と意味役割には相互に依存関係がある [渡邉 10] ことから，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に各動詞に対して語義ラベルを付与した．具体的には，動詞項構造シソーラスでも用いられている「基本語データベース:Lexeed」[笠原 04] の語義番号を付与した．Lexeed では 27,934 語の見出し語に対して，45,691 語義（1 語あたり平均 1.63 語義）付与されている．慣用句の一部として用いられている場合と Lexeed には無い語義で用いられている場合には，それぞれ「慣用句」と「新語義」の特殊ラベルを付与した．

NAIST テキストコーパスにおいて動詞は 21,706 種類，延べ 109,055 回出現するが，11 回〜15 回出現する動詞全 158 種，16 回出現する動詞の一部 14 種，81 回出現する動詞の一部 1 種，計 172 種 2,338 個の動詞の語義とその項の意味役割をアノテートした．

作成したデータの分析

4.1 統計

語義と意味役割の付与数

Lexeed では 173 動詞に対して，延べ 590 語義が定義されているが，今回付与した語義種類は 323 種類であった．

総意味役割付与数は 4,398 個である．表 3 にその内訳を示した．上位 3 グループの付与数は全体の 80%以上を占める．このような付与数の偏りは，付与対象を必須格のガ・ヲ・ニ格の項に限定したことが理由として考える．例えば，カラ格を付与対象にすれば，「場所」

[7]http://vsearch.cl.cs.okayama-u.ac.jp/semanticrole.php

[8]http://vsearch.cl.cs.okayama-u.ac.jp/

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)

表 4: 各格の項に付与した意味役割の頻度上位 6 種

格	意味役割の頻度上位 6 種	カバー率
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)	94.0%
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)	87.5%
ニ格 (605)	着点 (178), 対象 (68), 対象:事態 (61), 着点?(43), 着点:場所 (35), 対象:人 (29)	68.4%

や「方向」の意味役割は増え，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と考える．

各格の意味役割の分布

表 4 に各格の意味役割の頻度上位 6 種類を示した．

ガ格項の意味役割の大半は，[竹内 08] で議論されているように，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」の 3 つが占めている.「対象:事態」は「コメ生産」や「新会派結成」といった事態性名詞が項となるときに付与されている．

ヲ格項の意味役割は行為の直接対象を表す意味役割は「対象」（とその細分類）が多い.「生成物」は「その行為の結果生じるもの」で「対象」とは区別される．この意味役割はヲ格項を持つ 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．また，NAIST テキストコーパスでは必須項であるとされているが，そうではないと判断された「削除」はヲ格で最も多かった．これは，自動詞と他動詞が交替したり，項が事態性名詞となるときに多く付与された．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かった．

ニ格項は他の格と比べて，頻度上位 6 種類のカバー率が低く，意味役割の偏りが最も小さかった．また,「方向」「期限」,「内容物」などニ格にしか現れなかった意味役割もいくつかあった．

表 5: アノテーションの一致率

		Precision	Recall	F-measure
語義		Accuracy: 0.68 (63/92)		
意味役割	ガ格	0.98 (59/60)	0.65 (59/91)	0.78
	ヲ格	0.83 (24/29)	0.41 (24/58)	0.55
	ニ格	0.36 (4/11)	0.25 (4/16)	0.30

4.2 一致率

一致率を調査するため，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

語義ラベルの一致率は 68%であった．一致しなかった 29 事例を要因別に分けると，慣用句か否かの見解の違いが 8 事例，新語義か否かの見解の違いが 6 事例，語義の認識の違いが 14 事例であった．

意味役割ラベルの一致率

意味役割の一致率の評価は，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して，F 値で行う．結果を表 5 に示す．

意味役割の一致率はガ格，ヲ格，ニ格の順に低下する．意味役割の分布のばらつきもこの順に大きくなることから，ラベルの選択候補が増えることが理由の 1 つとして考えられる．

アノテーションの問題点と対応策

ここでは，アノテーションの際に判断が難しかった場合について説明し，その対応策を述べる．

5.1 語義付与の問題点

比喩的表現

比喩的な表現は，動詞と項だけでなく文脈を考慮する必要があるため，判断が難しかった．例えば，

風に 吹かれ るままに

の「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」である．一方，

運命の風に 吹かれ るままに

では，“運命の” という修飾語があるため，実際の物理現象の「風」が起きたわけではなく，比喩的表現として用いられてことから，ある作業者は「吹く」に「新語義」ラベルを付与した．ところが，比喩的に用いられているかの判断には作業者間での揺れがあった．そのため，比喩的な表現の可能性がある場合は，最も近い語義を選択した上でさらに「比喩的表現」ラベルを付与するのが適切だと考える．

5.2 意味役割付与の問題点

自動詞の使役形と他動詞の混同

動詞には「浴びる」と「浴びせる」のように自動詞と他動詞が対になっているものが多くある．また「浴

びさせる」のように自動詞は使役と結合可能である．例えば，

Aが Bに Cを浴びせた

Aが Bに Cを浴びさせた

では，前者は「浴びる動作」を Aが主体になって行うのに対して，後者は「浴びる動作」は Bが主体となって行う．ところが，今回のアノテーションではそれらが酷似していることから，自動詞の使役形と他動詞を混同してしまい，「浴びせる」の基本形は「浴びる」であるとして項の認定と意味役割の誤ったアノテーションが行われていた．そのため修正作業が必要であるが，その際にはそれぞれの動詞の自動詞であるか他動詞であるかをはじめに認識してから作業を行う必要がある．

意思性の有無

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

香港は . . . 経済発展を遂げた。

の「香港」が，意思をもって経済発展を「遂げた」のか，意思はなく偶然経済発展を「遂げた」のかは，この文からは判別できない．そのため，「経験者?」や「動作主?」といった意味役割ラベルは，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のが妥当である．

おわりに

本研究では，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．これにより，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．さらに，実テキストへのアノテーションを行ったことで，従来議論されてこなかった比喩的表現等の問題が明らかになった．今後の課題は，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ことと，必須格以外への意味役割のアノテーションを行うことである．

参考文献

[Baker 98] Baker, C. F., Fillmore, C. J., and Lowe, J. B.: The Berkeley FrameNet Project, in *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics*, Vol. 1, p. 86 (1998)

[Fillmore 82] Fillmore, C. J.: Frame semantics, in *Linguistics in the Morning Calm*, pp. 111–137 (1982)

[Jackendoff 90] Jackendoff, R.: *Semantic Structures*, The MIT Press (1990)

[Palmer 05] Palmer, M. and Kingsbury, P.: The Proposition Bank: An Annotated Corpus of Semantic Roles, *Computational Linguistics*, Vol. 31, No. 1, pp. 71–106 (2005)

[Surdeanu 03] Surdeanu, M., Harabagiu, S., Williams, J., and Aarseth, P.: Using predicate-argument structures for information extraction, in *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, Vol. 1, pp. 8–15 (2003)

[Taira 10] Taira, H., Fujita, S., and Nagata, M.: Predicate Argument Structure Analysis Using Transformation-based Learning, in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 162–167 (2010)

[Wu 09] Wu, D. and Fung, P.: Can Semantic Role Labeling Improve SMT?, in *Proceedings of the 13th Annual Conference of the European Association for Machine Translation*, pp. 218–225 (2009)

[飯田 10] 飯田龍, 小町守, 井之上直也, 乾健太郎, 松本裕治：述語項構造と照応関係のアノテーション: NAIST テキストコーパス構築の経験から, 自然言語処理, Vol. 17, No. 2, pp. 25–50 (2010)

[大石 95] 大石亨, 松本裕治：格パターン分析に基づく動詞の語彙知識獲得, 情報処理学会論文誌, Vol. 36, No. 11, pp. 2597–2610 (1995)

[小原 05] 小原京子, 大堀壽夫, 鈴木亮子, 藤井聖子, 斎藤博昭, 石崎俊：日本語フレームネット：意味タグ付きコーパスの試み, 言語処理学会第 11 回年次大会予稿集, pp. 1225–1228 (2005)

[小原 11] 小原京子：日本語フレームネットの全文テキストアノテーション: BCCWJ への意味フレーム付与の試み, 言語処理学会 第 17 回年次大会予稿集, pp. 703–704 (2011)

[影山 96] 影山太郎：動詞意味論 -言語と認知の接点-, くろしお出版 (1996)

[河原 07] 河原大輔, 黒橋禎夫：自動構築した大規模格フレームに基づく構文・格解析の統合的確率モデル, 自然言語処理, Vol. 14, No. 4, pp. 67–81 (2007)

[笠原 04] 笠原要, 佐藤浩史, Bond, F., 田中貴秋, 藤田早苗, 金杉友子, 天野成昭：「基本語意味データベース:Lexeed」の構築, 情報処理学会第 159 回自然言語処理研究会予稿集, pp. 75–82 (2004)

[竹内 06] 竹内孔一, 乾健太郎, 藤田篤：語彙概念構造に基づく日本語動詞の統語・意味特性の記述, レキシコンフォーラム No.2, pp. 85–120 (2006)

[竹内 08] 竹内孔一, 下村拓也：動詞語義を推定するための語義付与コーパスの作成, 言語処理学会 第 14 回年次大会予稿集, pp. 273–276 (2008)

[竹内 10] 竹内孔一, 土山傑, 守屋将人, 森安祐樹：類似した動作や状況を検索するための意味役割及び動詞語義付与システムの構築, 電子情報通信学会技術研究報告 言語理解とコミュニケーション研究会, pp. 1–6 (2010)

[竹内 11] 竹内孔一：動詞項構造シソーラスの構築, 人工知能学会 第 25 回全国大会予稿集, No. 3H2-OS3-5, pp. 1–4 (2011)

[橋田 05] 橋田浩一：GDA 日本語アノテーションマニュアル, http://i-content.org/gda/tagman.html (2005)

[肥塚 07] 肥塚真輔, 岡本紘幸, 斎藤博昭, 小原京子：日本語フレームネットに基づく意味役割推定, 自然言語処理, Vol. 14, No. 1, pp. 43–66 (2007)

[吉川 10] 吉川克正, 浅原正幸, 松本裕治：Markov Logic による日本語述語項構造解析, 情報処理学会 第 199 回自然言語処理研究会予稿集, 第 199 巻, pp. 5:1–7, 社団法人情報処理学会 (2010)

[渡邉 10] 渡邉陽太郎, 浅原正幸, 松本裕治：述語語義と意味役割の結合学習のための構造予測モデル, 人工知能学会論文誌, Vol. 25, No. 2, pp. 252–261 (2010)