

述語項構造解析のための格フレーム辞書構築の検討

田中 貴秋

永田 昌明

NTT コミュニケーション科学基礎研究所

{tanaka.takaaki, nagata.masaaki}@lab.ntt.co.jp

1 はじめに

統語解析の精度向上に伴い、述語と項の関係を捉える述語項構造解析に関する研究が盛んに行われている。日本語においては、言語資源としてNAISTテキストコーパス [1] の存在意義が大きい。述語項構造を意味構造からの写像と捉え、意味構造を復元する情報を獲得しようとする、付与されている項の情報が表層格「が」「を」「に」に限定されていること、述語の語義の区別がされていないこと、の2点において十分な情報を持っているとは言えない。

意味構造や述語の語義のような詳細な情報を含めて述語項構造情報をコーパス全体にわたって一貫性を保持して付与しようとするならば、各述語の語義ごとにそれぞれの項構造情報を列挙したもの（格フレーム [2]）が必要となる。英語においては、FrameNet [3] や PropBank [4] が各述語の語義ごとに、意味役割を含む項構造情報を辞書として持っている。一方、日本語では、同様の規模、品質を持つ言語資源が十分に整備されている状況にない。本稿では、このように述語項構造解析や意味構造解析を行う基盤の情報として、格フレーム情報を収録した辞書を構築する方法を検討した内容について述べる。以下、本稿では基本的な述語である動詞のみを対象とし、その他の述語については対象外とする。

2 格フレーム

2.1 定義

本稿では、述語項構造解析を行う基盤の情報として、格フレームを以下のように定義する。

- 動詞の語義ごとに、項構造情報を列挙する
- 項構造情報として、項ごとに格と意味役割と選択制限の情報をもつ

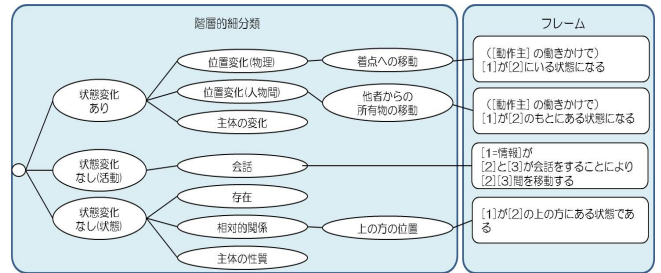


図 1: 動詞項構造シソーラス

同じ表層形を持つ動詞であっても、基本的に語義によって異なる項構造をもつと考えられるので、格フレームは各語義ごとに定義する。逆に類似した語義を持つ動詞では、類似した格フレームを持つ可能性がある。

項構造の情報としての意味役割は、動詞に対する項の役割を表し、複数の動詞の語義の間で、同様の役割を持つ項を識別するために付与する情報である (e.g. 動作主, 対象)。しかし、意味役割の種類には統一した結論がなく、各言語資源ごとに30種類程度から1,000種類以上まで様々な形で定義されている。

また、選択制限は各項の取り得る条件を表したもので、主に項になり得る名詞句の意味クラスを記述する。WordNet [5] や日本語語彙大系 [6] 等のシソーラスで定義された意味クラスで定義することが多いが、動詞によって各項の意味クラスの観点が異なるので、単一のシソーラスのみで選択制限を記述するのは難しい。

2.2 課題と方針

一貫性 類似した意味内容を持ち、類似した統語構造を持つ格フレーム間で、一貫した記述内容となる必要がある。また、前節でも述べたように、格フレーム内で動詞に対する項の意味役割を適切に定義する必要がある。

<< 焼く₁ >>
 (人) が
 (魚)|<魚介類> を

<< 焼く₂ >>
 (主体)|<機械> が|で
 (場所)|<具体物>|<場> を

<< 焼く₃ >>
 (人) が
 (パン)|<菓子>|<瓶・壺・盆> を

<< 焼く₄ >>
 (人) が
 (肉)|<卵> を

図 2: 「焼く」の格パターン(選択制限)の例

網羅性 国語辞書等で定義されている動詞の語義に基づき、シソーラスの意味クラスを用いて選択制限を記述した格フレーム辞書は人間の可読性が高い一方で、網羅性が問題になる。網羅性を高めるためには、河原らの提案した方法 [7] 等により、Web 等の大規模文書からの格フレームを自動獲得することが有効であるが、獲得された格フレームと動詞の語義との対応付けが課題となる。

そこで本稿では、動詞の語義を分類して体系化し、語義の分類単位で共通的な意味構造記述を行うことで一貫性を確保する。意味役割の定義はその共通的な意味構造記述の中で定義する。

また選択制限については、シソーラス上の意味クラスにより記述した可読性の高い格フレームの記述をベースとし、コーパス等からの用例の追加で補完することにより、記述の柔軟性を確保する方針とする。

3 動詞シソーラスと格パターン辞書に基づく格フレーム辞書構築

前節に述べたように、以下の方針に基づいて格フレーム辞書を構築する。

- 動詞の語義の分類と体系化
- 語義の分類単位で共通した意味構造の記述
- 意味クラスによる選択制限情報の記述と用例による補完

動詞の語義分類と分類単位ごとの意味構造記述については、竹内らが構築している動詞項構造シソーラス

[8]¹を、選択制限情報の付与については、日本語語彙大系を利用する。

3.1 動詞項構造シソーラス

動詞の語義を分類、抽象化して概念と定義し、この概念をノードとしてシソーラスを構築したものである。図 1 に示すように、このシソーラスの第一層には、「状態変化あり」「状態変化なし(活動)」「状態変化なし(状態)」の 3 分類があり、以下最大 5 階層²、約 700 分類(概念)から構成されている。各概念には、「構造的意味記述」という項構造を詳細化した形で、意味の記述がなされており、これにより項同士の意味的な関係が表現されている。この意味記述は概ね、動作主、様態、結果状態という 3 要素から構成されている。また、この記述には項の役割として約 70 種類のラベルが定義されており、これを格フレーム内の項の意味役割に利用する。

各ノード(概念)のインスタンスである動詞の語義は、「構造的意味記述」による項構造を共有し、各インスタンスごとに用例および、各項と意味役割の対応関係が記述されている。このため、類似した意味構造を持つ語義同士で意味役割等の一貫性を保ちやすく、格フレームの拡張は、このシソーラスの性質を利用して行う。

3.2 日本語語彙大系

「構文体系」辞書では、日本語から英語への訳し分けを目的として、各用言の取り得る格のパターンを名詞の意味クラス³による選択制限とともに記述している。意味クラスは階層的に配置された約 2,800 クラスからなり、約 6,000 の用言について約 14,000 の格パターンを収録している。図 2 は、収録されている格パターンの例である。「焼く」は翻訳のための 4 つの格パターンを持ち、それぞれのパターンは項の選択制限と格助詞の情報からなっている。<> で囲まれたラベルは、項となる単語の意味クラスを表している。訳し分けのための情報のため、日本語の中では同一の語義と見なせるものも別の格パターンとして表されている場合もある(「焼く₁」“grill”と「焼く₄」“roast”)。語義および格フレームの粒度としては、動詞項構造シソーラス上のノードを基本単位とし、同一ノード上に位置する同一単語の格フレームは一つに集約し、区別する必要があるが、

¹http://cl.cs.okayama-u.ac.jp/rsc/data/ から入手可能。

²階層数は特に制限されていない

³一般名詞意味属性と呼ぶ

<p>概念: 【対象の変化:性質の変化:化学的反応・変化】 構造的意味記述: ([動作主]の働きかけで) ([対象]が化学的に変化した状態になる)</p> <p>instance: 焼く₄ φ/[動作主] が 〈人〉 肉/[対象] を 〈肉〉 〈卵〉</p> <p>instance: 茹でる₁ φ/[動作主] が 〈人〉 卵/[対象] を 〈具体物〉</p>
<p>概念: 【生成・消滅:生成(物理):生成】 構造的意味記述: ([動作主]の働きかけで) ([生成物]が作られて) ([生成物]の存在する状態になる)</p> <p>instance: 焼く₃ φ/[動作主] が 〈人〉 器/[生成物] を 〈パン〉 〈瓶・壺・盆〉 窯/[道具] で -</p> <p>instance: 刺す₄ φ/[動作主] が 〈人〉 雑巾/[生成物] を 〈布〉 〈家庭用具〉</p>

図 3: 動詞項構造シソーラスの概念ノードと格フレームの対応例

下位のノードを作成して各ノードに格フレームを配置する方針とする。

3.3 格フレーム辞書の構築

動詞項構造シソーラスの各ノードのインスタンスである動詞の語義を単位として格フレームを構築する。各動詞の語義ごとに、用例と構造的意味記述に基づいて各項と意味役割との対応がつけられているので、これを格フレームのベースとする。各項には選択制限の情報がないので、日本語語彙大系から一致する格パターンを探し付与する。

動詞「焼く」の格フレームを作成する場合を例として説明する。図 3 は、「焼く」の語義がインスタンスとなっている動詞項構造シソーラスのノードの例である。「焼く」の語義がインスタンスとなっている概念には、【対象の変化:性質の変化:化学的反応・変化】(以下【化学的反応・変化】)と【生成・消滅:生成(物理):生成】(以下【生成】)がある。⁴

【化学的反応・変化】の方は、構造的意味記述は[動作主]と[対象]の意味役割を持つ項を用いて、「[動作主]の働きかけ」により「[対象]が化学的に変化する」結果状態を生じることを示している。この意味役割と

⁴動詞項構造シソーラス上の概念を、最上位階層のノードから順に「:」で連結し【】で囲んで表す。いずれもシソーラス第 1 層の「状態変化あり」を省略している。

用例から、この語義の格フレームは「[動作主]が肉/[対象]を焼く」という構造になっている。次に、用例を日本語語彙大系の格パターンと照合すると、「を格」に「肉」が入るのは ≪ 焼く₄ ≫ の格パターンであるので、この対応付けにより選択制限の情報を付加することができる。

また、【生成】の概念のインスタンスになっている方も同様に ≪ 焼く₃ ≫ の格パターンが適合する。用例では、意味役割 [道具] を持つ「で格」が入っているが、構造的意味記述に含まれていないことから、概念のコア部分でないと考えて、随意的な要素と扱う。

格フレームは、概念ノードのインスタンスである動詞の語義に対応づけられるので、同じ概念に属する格フレーム同士は、同様の意味構造を共有していると考えられる。例では、≪ 焼く₄ ≫ と ≪ 茹でる₁ ≫ はともに、「〈人〉/[動作主]が〈具体物〉/[対象]を」という構造になっている。

同じ概念に属する格フレーム間での類似性、共通性を維持するように構築することで、類似の動詞の語義に関する記述の一貫性を保つことができる。

3.4 用例を用いた格フレーム辞書の拡張

前節で述べた方法で構築した格フレームは、項構造の概念関係を意識しつつ構築することができるが、実データの解析等に適用する場合には、選択制限の記述が不十分であったり、格フレームの不足等により適合する格フレームが存在しない場合がある。本節では、人手で構築した格フレームをベースとして、コーパスの用例を取り込むことにより、網羅性を向上させていく方法について述べる。

基本的な考え方は、収集した用例を格フレームの選択制限と照合し、一つの格フレームに適合するもののみを、各格フレームの正例として分類モデルを構築し、選択制限のみでは選別できない用例(適合する格フレームがない、または複数の格フレームに適合する)を、各格フレームに分類する。

以下に処理の概要を記す。

1. コーパスから対象とする動詞を中心とする述語項構造の用例を収集する
2. 収集した述語項構造を、既存の格フレーム辞書を用いて、選択制限に適合するものを振り分け、各格フレームの用例とする。どの格フレームの条件にも適合しないもの、複数の格フレームの条件に適合するものを、分類すべき述語項構造の用例としてストックする。

パン-を	1606
オープン-で	1100
ケーキ-を	830
手-を	795
世話-を	786
肉-を	711
CD-に *	591
-	
CD-に; ファイルを *	13
CD-に; データを *	9

図 4: 「焼く」の項とコーパス中出現頻度の例: * は適合する格フレームが存在しないものを示す

- 各格フレームの用例を用いて、格フレームの分類モデルを構築する
- 分類すべき述語項構造の用例を、分類モデルを用いて格フレームに分類する (ある閾値によりどれにも分類できなかった述語項構造は、新たな格フレームが必要であると判断する)
- 新たに分類された用例を用いて選択制限を更新し、2へ戻る。

結果として、新たに分類された述語項構造の用例を用いて選択制限を修正していく。また、新たな格フレームが必要であると判断された述語項構造の用例については、動詞項構造ソーラス上で、類似した構造的意味を持ちかつ統語的に近い候補を見つけて、新たな格フレームを作成する。

図 4 に、Web からダウンロードしたデータ (約 180 億語) から抽出した「焼く」の項の例があるが、⁵。この例では、「DVD-に」「CD-に」などの項が適合する格フレームがない。新たな格フレーム (語義) を追加する際には、既存の同義語との間で整合性を取る必要がある。たとえば、「データを CD に焼く」の「焼く」の格フレームを追加する際には、言い換え可能な同義語「記録する」があるため、「記録する」の語義が属しているノード【位置変化 (情報) (媒体の変換)】上に配置し、用例を「データ/[対象]-を CD/[着点]-に焼く」として蓄積する (図 5)。

4 おわりに

本稿では、一貫性と網羅性を確保した格フレームの構築を目的として、動詞の語義の概念の体系として動詞項構造ソーラスを、選択制限のベースとして日本語語彙大系を用いた方法について述べた。今後は、格

⁵統語解析器の解析結果から動詞に係る文節の組み合わせを抽出している。

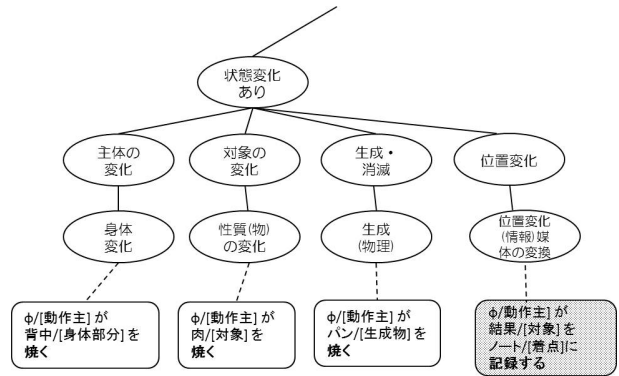


図 5: 動詞「焼く」の語義、動詞項構造ソーラス、格フレームの関係

フレーム辞書を構築し、コーパスへのアノテーションや解析等への利用により評価を行う予定である。

参考文献

- Iida, R., Komachi, M., Inui, K., and Matsumoto, Y.: *Annotating a Japanese text corpus with predicate-argument and coreference relations*. In *Proceedings of Linguistic Annotation Workshop, Association for Computational Linguistic Annotation Workshop (ACL 2007)*, pp. 132-139, (2007).
- Fillmore, C. J.: "A private history of the concept 'frame'," Dirven, Rene and Radden, Gunter (Eds.), *Concept of Case*, (1987).
- Baker, C. F., Fillmore, C. J., and Lowe, J. B.: The Berkeley FrameNet project, In *Proceedings of 36th Annual Meeting of the Association for Computational Linguistics (ACL 1998)*, pp.86-90, (1998).
- Palmer, M., Gildea, D., Kingsbury, P.: *The Proposition Bank: An Annotated Corpus of Semantic Roles*. *Computational Linguistics*, Vol. 31, No. 1, pp.71-106, (2005).
- Fellbaum, C.: *WordNet an Electronic Lexical Database*. MIT Press (1998).
- 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦: 日本語語彙大系, 岩波書店, (1997).
- 河原大輔, 黒橋禎夫: 格フレーム辞書の漸次的自動構築, 自然言語処理, Vol.12, No.2, pp.109-131, (2005).
- 竹内孔一: 動詞項構造ソーラスの構築, 人工知能学会全国大会, 3H2-OS3-5, (2011).